DOCUMENT RESUME

ED 317 611 TM 014 725

AUTHOR	Shields, Patrick M.; And Others
TITLE	Guidance to States on Evaluating the Effectiveness of the Chapter 2 Program.
INSTITUTION	SRI International, Menlo Park, Calif.
SPONS AGENCY	Department of Education, Washington, DC. Office of Planning, Budget, and Evaluation.
PUB DATE	Oct 89
CONTRACT	LC890160
NOTE	100p.
AVAILABLE FROM	SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025.
PUB TYPE	Guides - Non-Classroom Use (055)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Compensatory Education; *Data Analysis; Data Collection; Educational Finance; Elementary Secondary Education; Evaluation Criteria; *Evaluation Methods; *Federal Programs; Federal State Relationship; Program Effectiveness; *Program Evaluation; *Research Design; State Agencies; *State Programs
IDENTIFIERS	*Education Consolidation Improvement Act Chapter 2

ABSTRACT

Guidance is provided for state Chapter 2 coordinators and evaluators in designing and implementing the congressionally mandated evaluation of the effectiveness of the Chapter 2 program. The first part of this paper reviews the congressional mandate and discusses the appropriate audience for evaluation reports. A schedule is outlined for conducting reporting and evaluation over 3 years. A framework is given for an evaluation design that includes a typology of Chapter 2-funded activities and a model of how these activities influence the educational system and student learning. Operational suggestions for staffing and funding are offered. The second part of this paper uses the typology found in the first part to discuss the evaluation of different types of activities, with emphasis on design, data collection, and data analysis. The third part provides more detailed guidance on specific design issues and discusses: (1) sampling strategies; (2) survey design and analysis; and (3) case studies. Seventeen tables, 8 figures, and a 10-item selected bibliography are included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED317611

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

A. BRENNO

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

GUIDANCE TO STATES ON EVALUATING THE EFFECTIVENESS OF THE CHAPTER 2 PROGRAM

OCTOBER 1989

Prepared by:

Patrick M. Shields
Michael S. Knapp
E. Deborah Jay

SRI International
Menlo Park, CA

U.S. ED Contract LC890160

U.S. DEPARTMENT OF EDUCATION • OFFICE OF PLANNING, BUDGET & EVALUATION

TM014925

BEST COPY AVAILABLE

CONTENTS

PART ONE: OVERVIEW OF APPROACH TO EVALUATING THE EFFECTIVENESS OF CHAPTER 2 1

I INTRODUCTION 3

The Audience and Expectations for the SEA Evaluations of Chapter 2 Program Effectiveness 4

Schedule 5

II A FRAMEWORK FOR DESIGNING AN EVALUATION OF CHAPTER 2 7

A Typology of Chapter 2-Sponsored Activities 8

An Overall Analytical Model of Chapter 2 Influences on Education and Student Learning 11

Using the Analytical Model to Think About Effectiveness . . 13

III PUTTING TOGETHER A STATE EVALUATION PLAN 15

Defining the Scope of State Evaluation Efforts 15

Specifying an Analytical Model for Each Type of Chapter 2-Sponsored Activity to Be Evaluated 16

Choosing an Appropriate Design and Data Collection Methods . 21

Adopting an Approach to Analyzing and Reporting Information 24

Gathering Sufficient Resources to Carry Out the Evaluation . 28

PART TWO: APPROACH TO EVALUATING EACH TYPE OF CHAPTER 2-SPONSORED ACTIVITY 35

IV EVALUATING THE EFFECTIVENESS OF CHAPTER 2 SUPPORT FOR PROFESSIONAL DEVELOPMENT 37

Defining and Targeting Professional Development Activities for Evaluation 37

Specifying an Analytical Model for Evaluating Professional Development Activities 39

CONTENTS (Continued)

Choosing Design and Data Collection Methods 42

Adopting an Approach to Data Analysis and Reporting 43

V EVALUATING THE EFFECTIVENESS OF THE ACQUISITION AND DEVELOPMENT OF MATERIALS AND COMPUTER EQUIPMENT 45

Defining and Selecting Material and Equipment Acquisition and Development Activities for Evaluation 45

Specifying an Analytical Model for Evaluating the Acquisition or Development of Material and Computer Equipment 47

Choosing Design and Data Collection Methods 50

Adopting an Approach to Data Analysis and Reporting 51

VI EVALUATING THE EFFECTIVENESS OF DIRECT SERVICES TO STUDENTS 53

Defining and Selecting Activities that Provide Direct Services to Students for Evaluation 53

Specifying an Analytical Model for Evaluating Activities that Provide Direct Services to Students 55

Choosing Design and Data Collection Methods 58

Adopting an Approach to Data Analysis and Reporting 59

VII EVALUATING THE EFFECTIVENESS OF SUPPORT FOR SCHOOLWIDE IMPROVEMENT, INCLUDING EFFECTIVE SCHOOLS PROGRAMS 61

Defining and Selecting School Improvement and Effective Schools Activities 61

Specifying an Analytical Model for Evaluating School Improvement and Effective Schools Programs 63

Choosing Design and Data Collection Methods 66

Adopting an Approach to Data Analysis and Reporting 67

CONTENTS (Continued)

VIII EVALUATING THE EFFECTIVENESS OF CHAPTER 2-SPONSORED ACTIVITIES DIRECTED AT AT-RISK STUDENTS 69

IX EVALUATING THE ACHIEVEMENT OF OTHER LEGISLATIVE GOALS UNDER CHAPTER 2 71

Evaluating the Contribution of Chapter 2 to Private School Students 71

Evaluating the Effectiveness of Chapter 2-Supported Demonstration Projects 72

PART THREE: OPERATIONAL SUGGESTIONS AND USEFUL REFERENCES 75

X SAMPLING STRATEGIES 77

Sampling Districts 78

School Samples 80

Sampling Teachers 81

The Issue of Survey Response Rates 82

The Issue of Weighting 83

XI SURVEY DESIGN AND ANALYSIS ISSUES 85

Instrument Development 85

Data Collection 91

Data Processing 95

Measurement Construction 95

Analysis 96

XII CASE STUDIES 99

Sampling Cases 99

CONTENTS (Continued)

Instrument Development 100

Preparing for Data Collection 101

Conducting District Visits 102

Data Analysis and Reporting 103

XIII SELECTED BIBLIOGRAPHY OF RESOURCES 105

LIST OF TABLES

Table 1. SEA Evaluation and Annual Report Schedule 6

Table 2. Typology of Chapter 2-Funded Activities 9

Table 3. Evaluation Questions and Indicators of Effectiveness for Leadership Training Programs for Principals 20

Table 4. Overview of Data Collection Mechanisms for Evaluating the Effects of Leadership Training Workshops 22

Table 5. Typology of Professional Development Activities 38

Table 6. Evaluation Questions and Indicators of Effectiveness for Professional Development Activities 41

Table 7. Overview of Data Collection Mechanisms for Professional Development Activities 42

Table 8. Typology of Material and Equipment Acquisition and Development Activities 46

Table 9. Evaluation Questions and Indicators of Effectiveness for the Acquisition and Development of Material and Computer Equipment 49

Table 10. Overview of Data Collection Mechanisms for Material and Computer Equipment Acquisition and Development 50

Table 11. Typology of Activities that Provide Direct Services to Students 54

Table 12. Evaluation Questions and Indicators of Effectiveness of Activities that Provide Direct Services to Students . . 57

Table 13. Overview of Data Collection Mechanisms for Activities that Provide Direct Services to Students 58

Table 14. Typology of Effective Schools Activities 62

Table 15. Evaluation Questions and Indicators of Effectiveness of Effective Schools Programs 65

Table 16. Overview of Data Collection Mechanisms for Effective Schools Programs 66

Table 17. Choosing Overall Sample Sizes for District Surveys 78

LIST OF FIGURES

Figure 1. An Analytical Model of Chapter 2 Influences on Education and Student Learning 12

Figure 2. Simple Model of Leadership Training Effects 17

Figure 3. Elaborated Model of Leadership Training Effects 17

Figure 4. An Analytical Model of the Effects of Professional Development Activities 39

Figure 5. An Analytical Model of the Effects of Material Acquisition and Development 47

Figure 6. An Analytical Model of the Effects of Activities that Provide Direct Services to Students 55

Figure 7. An Analytical Model of the Effects of Effective School Programs 63

Figure 8. Sampling Strategy for Various Data Collection Methods . . . 82

PART ONE: OVERVIEW OF APPROACH TO EVALUATING THE EFFECTIVENESS OF CHAPTER 2

The purpose of this document is to provide state Chapter 2 coordinators and evaluators formal guidance in designing and implementing the congressionally mandated evaluation of the Chapter 2 program's effectiveness.

In this first part of the Guidance, we review the congressional mandate for state evaluations. We also discuss the appropriate audience for the evaluation reports and outline a schedule for carrying out both reporting and evaluation over the next three years.

We continue in this part by providing a framework for the design of Chapter 2 evaluations. The framework includes a typology of Chapter 2-funded activities and an analytical model of how these activities influence the educational system and student learning.

In the final section of Part One, we offer a set of operational suggestions for garnering sufficient resources for staffing and funding the evaluation efforts.

I INTRODUCTION

The purpose of this document is to provide state Chapter 2 coordinators and evaluators formal guidance in designing and implementing the congressionally mandated evaluation of the the Chapter 2 program's effectiveness.

The 1988 Hawkins-Stafford Elementary and Secondary School Improvement Act replaced Chapter 2 of the Education Consolidation and Improvement Act with Chapter 2 of the Title I of the Elementary and Secondary Education Act of 1965, as amended. Entitled "Federal, State, and Local Partnership for Educational Improvement," the new legislation continues to provide financial assistance to state education agencies and through them to local districts. Congress' decision to retain the basic program structure, which allows state and local agencies significant latitude in the use of those funds, reflects its finding that the Chapter 2 program has generally lived up to its expectations. Specifically, Congress found that the program *"has been successful in achieving the goals of increasing local flexibility, reducing administrative burden, providing services for private school students, encouraging innovation, and contributing to the improvement of elementary and secondary education programs."* 20 U.S.C. 2911(a).

At the same time, however, Congress responded to criticism that the old Chapter 2 program was relatively unfocused in many local settings and that there was insufficient accountability. Accordingly, while the new legislation retains its original framework, it identifies specific purposes for which funds should be targeted. Moreover, Congress built in a series of accountability mechanisms. Specifically, the law requires:

- Local education agencies (LEAs) to report annually to the State education agency (SEA) on their use of Chapter 2 funds.
- SEAs to report annually to the Secretary data on the use of Chapter 2 funds by the SEA and the LEAs, including the types of services provided and the children receiving those services.

- SEAs to evaluate the effectiveness of state and local Chapter 2 programs, due to the Secretary in 1992.

- The Secretary to submit a report to Congress summarizing the state evaluations in order to provide a national overview of the uses of funds and effectiveness of Chapter 2 programs. 20 U.S.C. 2932(a)(6)-(7), 2943(a)(4), 2973.

State agencies, then, are required to submit (1) annual reports, which provide descriptive data on the uses of Chapter 2 funds, and (2) a formal evaluation, which assesses the effectiveness of Chapter 2-funded services. In order to assist SEAs, Congress also called on the Secretary to provide written guidance for state agencies on effective ways to carry out these reporting and evaluation functions. This document constitutes the Secretary's Written Guidance for evaluating the effectiveness of Chapter 2-sponsored activities. The Secretary will disseminate a model reporting form in a separate document.

The Audience and Expectations for the SEA Evaluations of Chapter 2 Program Effectiveness

The U.S. Department of Education should be considered the primary audience for the SEA evaluation.

The central purpose of the SEA Chapter 2 evaluations is to provide Congress with information on the educational impact of Chapter 2-funded activities as it considers the reauthorization of the legislation in FY 1993. Because the Secretary of Education holds responsibility for summarizing the results of the SEA evaluations in a formal report to Congress, the Department of Education is the key audience for these evaluations. At the same time, state agencies should design evaluations that will provide them with data useful for their program reporting and improvement efforts.

Although the goals of each state's evaluation will reflect its specific program efforts, SEA staff should be aware that the overarching federal goal is to assess the impact of Chapter 2 on the improvement of education. SEA staff should, then, design and implement their evaluations to respond to the following three central questions:

(1) What educational services are funded by the Chapter 2 program?

(2) What students and staff are affected by those services?

(3) What are the educational impacts of those services (on schools' curriculum and instruction, on school staff, and on students)?

Responding to this third question regarding educational impacts is imperative.

> **Congressional and Administration intent is clear: state agencies are supposed to evaluate the effectiveness of their efforts, to link Chapter 2-funded services to student outcomes or some other reasonable education goal.**

Schedule

SEA preparation of both annual reports of Chapter 2-funded services and the evaluation of the effects of those services have to follow a tight timeline dictated by the schedule of congressional appropriations and reauthorization. Table 1 provides an overview of the schedule. During the 1989-90 school year (SY), SEA staff should develop mechanisms for receiving descriptive data from their LEAs for the annual reports and begin to collect data for that year's report, due to ED in December of 1990. SEA staff should also design and field test their evaluation of the effectiveness of the program by the end of the 1989-90 school year. In SY 1990-91, 1991-92, and 1992-93, SEA staff should collect data for their annual reports, which will be due each December. Submission of reports on the previous school year each December affords ED sufficient time to deliver summaries to Congress before the annual appropriation process.

Data collection for the evaluation of program effectiveness should take place during SY 1990-91. Note that states will only be expected to have evaluative data for this school year, although they should feel free to supplement it with data from previous years, when available. The evaluation will be due to ED by March of 1992. ED in turn has to report to Congress no later than October of 1992. It is at this point that Congress will be considering the reauthorization of the Chapter 2 legislation.

Table 1

SEA EVALUATION AND ANNUAL REPORT SCHEDULE

School Year	SEA Annual Reports	SEA Evaluations of Effectiveness	Department of Education Summary
1989-90	Establish LEA-SEA collection mechanism Collect data on SY 89-90	Design and pretest	--
1990-91	Collect data on SY 90-91 Report to ED (12/90) on SY 89-90	Implement evaluation	Summary of SY 89-90 annual report to Congress (4/91)
1991-92	Collect data on SY 91-92 Report to ED (12/91) on SY 90-91	Report to ED 3/92	Summary to Congress of SEA evaluations (10/92) Summary of SY 90-91 annual reports to Congress (4/92)
1992-93	Collect data on SY 92-93 Report to ED (12/92) on SY 91-92	--	Summary of SY 91-92 annual reports to Congress (4/93)

II A FRAMEWORK FOR DESIGNING AN EVALUATION OF CHAPTER 2

An effective evaluation plan should rest on a framework that: (1) clearly defines the activities to be studied and (2) specifies how these activities are expected to affect educational outcomes.

In this section, we take on these tasks, first by presenting a typology of Chapter 2-sponsored investments--sets of activities that can be evaluated together. We then develop a generic framework for thinking through how the different types of activities relate to one another and may affect student outcomes. Finally, we discuss how evaluators can use the framework to think about effectiveness.

In our discussion we bear in mind the difficulties inherent in evaluating Chapter 2-funded activities, due to the facts that:

- Chapter 2 is not a single, well defined innovation. The law's purposeful flexibility means that educational efforts funded by Chapter 2 vary considerably across states and local agencies.
- The legislation is not targeted on a single population; for example, some services are focused on gifted and talented students while others target low achievers.
- Chapter 2 funds often constitute a marginal addition to local efforts that are already underway--helping a school increase its planned purchase of computers, for example.
- The amount of Chapter 2 funds are typically small in relation to a school's or district's other financial resources.

Assessments of Chapter 2 then might involve judgments of the marginal effect of a relatively small amount of funds on a wide variety of programs aimed at different audiences. In these cases, establishing the relationship between the input of Chapter 2 dollars into the system and student achievement or some other system outcome can be quite difficult.

A Typology of Chapter 2-Sponsored Activities

In order to plan a manageable evaluation effort, evaluators have to be able to categorize Chapter 2-sponsored activities in some meaningful way. The logical starting point for the development of these categories is the set of targeted purposes outlined in the Chapter 2 legislation, which are also the organizing structure for each state's annual report to the U.S. Department of Education regarding Chapter 2-funded services:

(1) Programs to meet the educational needs of students at risk.

(2) Programs for the acquisition and use of instructional materials.

(3) Innovative programs designed to carry out schoolwide improvements, including effective schools programs.

(4) Programs of training and professional development.

(5) Programs designed to enhance personal excellence of students and student achievement.

(6) Other innovative projects to enhance the educational program and climate of the school. 20 U.S.C. 2941(b).

These targeted purposes are not mutually exclusive, however; nor are they strictly comparable. For example, some refer to targets of activities (at-risk students), while others address specific mechanisms for implementing improvements (professional development). Some of these categories, then, do not work well for evaluation purposes. One more step is needed to create categories that allow for meaningful comparisons across activities.

Table 2 provides a matrix that accomplishes this task. This simple matrix allows the evaluator to categorize Chapter 2-supported efforts first by the mechanism through which they are designed to meet their goals: whether they involve the acquisition and development of materials and computer equipment, the training of staff, or direct services to students (we also provide an "other" category). It then calls for the evaluator to classify the activity along two other dimensions: whether it is targeted on at-risk students and whether it is part of a broader schoolwide improvement effort.

Table 2

TYPOLOGY OF CHAPTER 2-FUNDED ACTIVITIES

	Part of Schoolwide Improvement Efforts? (including effective schools efforts)				
	Yes		No		
	Targeted on At-Risk Students?		Targeted on At-Risk Students?		
	Yes	No	Yes	No	
Professional Development					(A)
Acquisition and Use of Materials and Computer Equipment					(B)
Direct Services to Students*					(C)
Other**					(D)
	(1)	(2)	(3)	(4)	

* Includes programs designed to enhance the personal excellence of students and other innovative projects designed to enhance the educational program and climate of the school when they involve direct services to students.

** Includes other innovative projects to enhance the educational program of the school which do not involve professional development, material acquisition, or direct services to students.

Although the cells of the matrix do not match the legislation's targeted purposes exactly, each purpose is reflected in one or more columns or rows:

(1) Programs targeted on at-risk students are found in Columns 1 and 3.

(2) Programs involving the acquisition and use of materials are found in Row A.

(3) Innovative schoolwide improvements (including effective schools programs) are found in Columns 1 and 2.

(4) Programs of training and professional development are found in Row B.

(5) Programs designed to enhance personal excellence and student achievement which involve direct services to students are found in Row C.

(6) Other innovative projects to enhance the educational program and climate of the school are found both in Row C (when they involve direct services to students) and Row D.

The matrix allows the evaluator to aggregate data across categories and to report data to the federal government by the Chapter 2 legislation's targeted purposes. This match-up between targeted purposes also enables states to link their evaluations with the annual reports. For example, one could easily group all material acquisition activities or all efforts targeted at at-risk students. The use of the matrix to categorize Chapter 2-sponsored activities for evaluation purposes is preferable to simple listing by targeted purpose because it provides conceptually coherent categories within which the evaluator can make meaningful comparisons across similar activities (e.g., all activities aimed at training teachers to address the needs of at-risk students). The matrix has two additional strengths:

- The categories are sufficiently broad that most Chapter 2-funded activities are accommodated easily.
- It can be used to classify both state- and locally administered activities. For use in categorizing state-administered activities, however, the "professional development" category would have to be broadened to include state technical assistance efforts, some of which involve strict staff development, others of which simply involve state professionals working with local educators to devise more effective educational strategies.

The inclusion of state-administered programs is important because these often comprise the most innovative uses of Chapter 2 dollars.

In presenting this matrix, we do not mean to imply that there are not other dimensions along which evaluators might want to categorize Chapter 2-

funded activities. For example, it might be useful to note the grade levels or the academic content area involved in a set of activities. For reporting evaluation results to the federal government, however, the three dimensions included in Table 2 are the most important.

An Overall Analytical Model of Chapter 2 Influences on Education and Student Learning

The second conceptual task facing the evaluator is developing an analytical model that portrays how different types of Chapter 2 activities relate to one another and influence student learning. Here we present a generic model; later in the guidance we provide more detailed models for the different types of activities.

As depicted in Figure 1, the model assumes that all activities originate in some facilitating effort on the part of the state agency. In most cases, this simply involves the passing on of 80% or more of the state's allocation to the local districts through formulas or competitive grants. In many cases, however, the state uses its 20% set-aside to fund staff development programs, to purchase computer equipment for schools, and to sponsor local school improvement efforts. At the LEA level, a large portion of local funds are expended on material acquisition (or development) and professional training. These in turn may feed into broader school improvement efforts, or they may be designed to affect student services directly. Some local funds are expended immediately on direct services to students (some of whom may be targeted as at-risk). Finally, at the end of this process, these various activities are expected to affect student performance.

For example, LEAs may utilize Chapter 2 funds to sponsor staff development programs for elementary school teachers in a particular curricular area (e.g., reading). Using this framework allows one to note that SEA funds to the local district (state facilitation) support teacher training (professional development) activities that are designed to improve the type of reading instruction students receive (direct services), which in turn are expected to improve student achievement (outcomes).

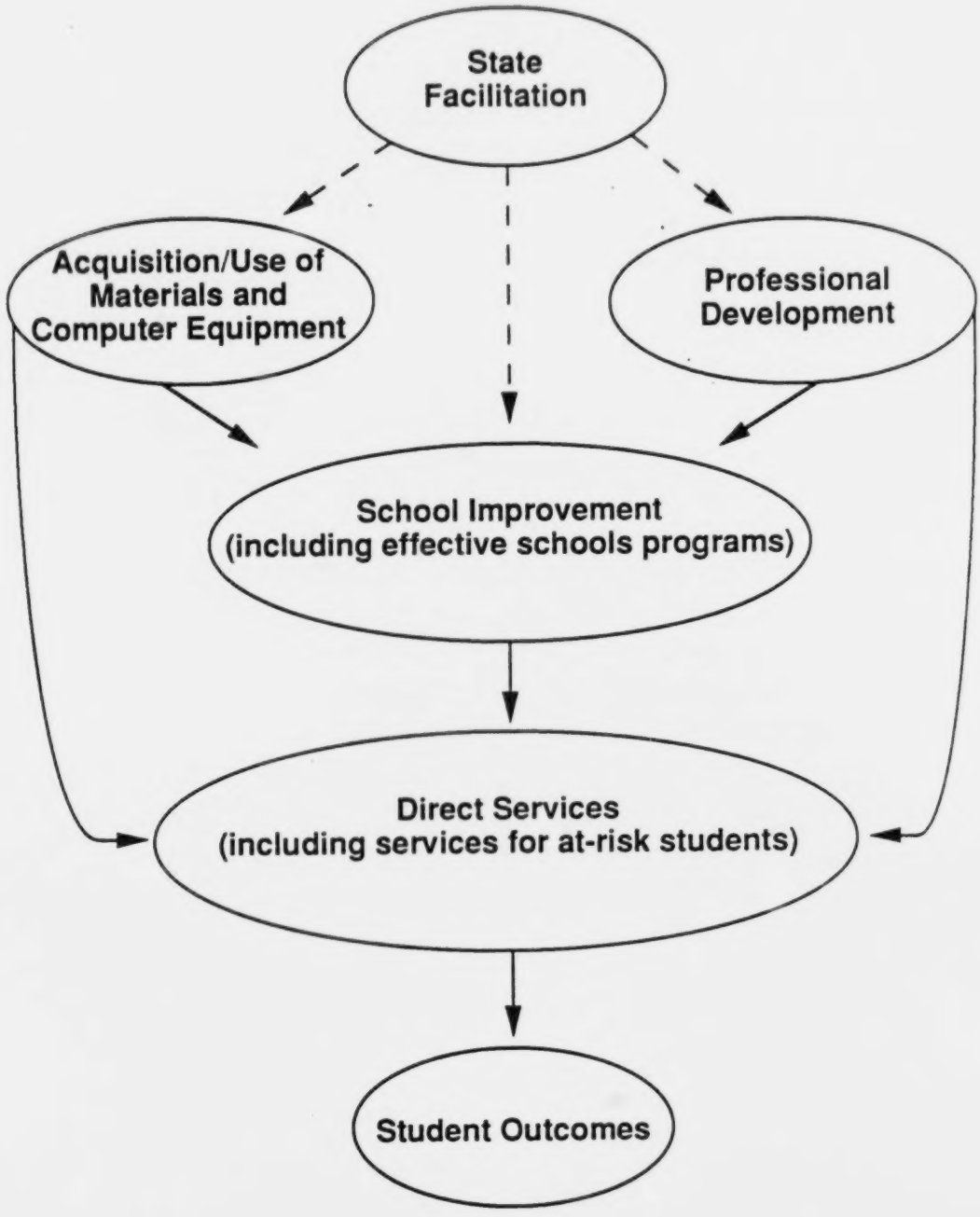

State Facilitation
Acquisition/Use of Materials and Computer Equipment
Professional Development
School Improvement (including effective schools programs)
Direct Services (including services for at-risk students)
Student Outcomes

FIGURE 1 AN ANALYTICAL MODEL OF CHAPTER 2 INFLUENCES ON EDUCATION AND STUDENT LEARNING

Using the Analytical Model to Think About Effectiveness

This generic model is not specific enough to be used to design an evaluation of particular activities. But it does highlight two factors common to Chapter 2 efforts that evaluators need to consider when attempting to assess program effectiveness.

- Chapter 2-funded activities are often interrelated (e.g., many school improvement programs turn on a combination of staff development and new materials).
- Many local Chapter 2-funded activities (and nearly all state-sponsored activities) do not directly affect the type of services students receive. Most frequently they affect professionals or materials most directly, which, in turn, will (or can) affect the education students receive.

The implication for evaluating effectiveness is simple:

> **One should not think of most Chapter 2-funded activities as a direct influence on student learning outcomes. Rather, Chapter 2-sponsored efforts must first be assessed in terms of their most immediate effects on a broad educational system.**

For example, the effectiveness of staff development programs may be measured initially by their effect on teachers' motivations, grasp of the discipline they teach, or ideas about how to teach it; subsequently, the effect of the training on teachers' classroom practice (and so on student learning experiences) can be assessed. Ultimately, it may be possible to assess the impact of the training on student learning. Within this framework, then, one can begin to design evaluation strategies that capture the immediate effects of Chapter 2-sponsored interventions and lay the groundwork for examining longer-term outcomes.

III PUTTING TOGETHER A STATE EVALUATION PLAN

An evaluation plan that will effectively guide evaluators through data collection and analysis should include the following:

(1) A definition of the scope of the state evaluation effort.

(2) An analytical model for each of the main types of Chapter 2-sponsored activity to be evaluated.

(3) Appropriate design and data collection methods.

(4) An approach to analyzing and reporting results (see sample evaluation report outline at end of section).

(5) Plans for garnering sufficient resources to carry out the evaluation.

In this section, we address each of these five points. In Part Two of the Guidance, which follows, we apply these principles to evaluation designs for each of the different types of Chapter 2-sponsored activities.

Defining the Scope of State Evaluation Efforts

The scope of evaluation may or may not include all activities funded under Chapter 2. The state may choose to focus on those targeted areas that receive most of the Chapter 2 funding, are most widespread, or most closely related to significant state needs. Operationally, these choices may mean that the state collects and reports data on the effectiveness of Chapter 2-sponsored activities in only some of the targeted areas.

Because evaluating the Chapter 2 program is such a complicated task and no state has the resources to carry out a thorough evaluation of each state

or local Chapter 2-sponsored activity, state evaluators have to make decisions about where to target limited assessment funds. Given knowledge of the range and type of Chapter 2-funded activities in the state (which can be categorized by the matrix presented in Table 2), states need to identify the specific targets of the evaluation and justify their selection.

Making choices among the wide variety of activities underway across the state requires the employment of reasonable criteria that justify a focus on one set of activities instead of others. States are free to use any reasonable criteria; however, the following three criteria seem especially relevant to federal audiences:

- The proportion of funds allocated to the activity.
- The proportion of LEAs conducting the activity.
- The extent to which the activity addresses what has been identified by the state as a significant need.

Thus, for example, a state could choose to target local material acquisition activities because nearly all LEAs in that state use funds for this purpose. Or the SEA may choose to look at high school dropout prevention programs because, although small in number, they are addressing a significant problem in the state.

Specifying an Analytical Model for Each Type of Chapter 2-Sponsored Activity to Be Evaluated

A well-conceived analytical model helps the evaluator to set realistic expectations for what effects can and cannot be measured, to develop specific evaluation questions, and to pinpoint potential indicators of effectiveness.

The second step in developing an evaluation plan involves an understanding of how each of the activities to be evaluated works: how it is expected to affect the educational system and result in student level effects.

Let us take, as an example, the evaluation of a set of locally sponsored leadership training programs for high school principals. In the absence of a well-developed analytical model, we are faced with the following frame for the evaluation (Figure 2), driven by the overall question about the effect of the training on student outcomes:

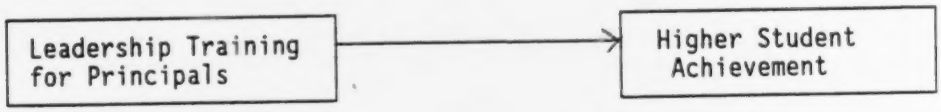


Figure 2

SIMPLE MODEL OF LEADERSHIP TRAINING EFFECTS

This frame suggests only one evaluation question ("What was the effect of leadership training on student achievement?") and one indicator of effectiveness (student achievement). It also suggests only two points at which to collect data (Which principals were trained? What was the change in average student test scores in their schools?). If one were to collect data at these two points and find higher test scores, it may be difficult to attribute them to the training; similarly, lack of rise in test scores might not indicate anything about the success of the training.

A better-developed model would look like this:

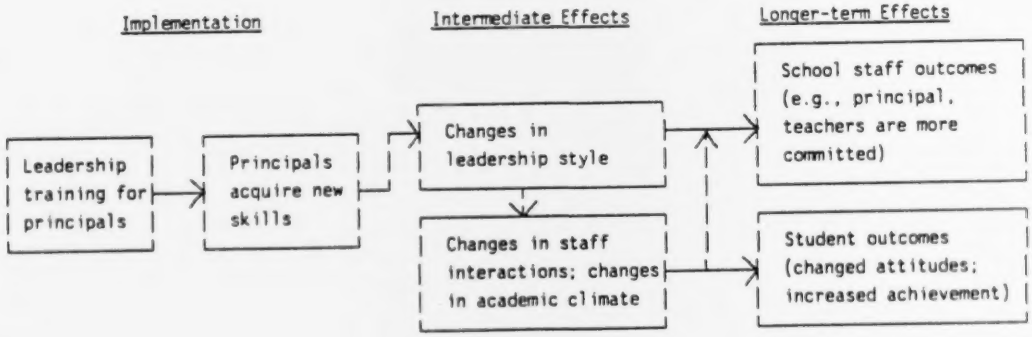

Implementation
Intermediate Effects
Longer-term Effects
Leadership training for principals
Principals acquire new skills
Changes in leadership style
Changes in staff interactions; changes in academic climate
School staff outcomes (e.g., principal, teachers are more committed)
Student outcomes (changed attitudes; increased achievement)

Figure 3

ELABORATED MODEL OF LEADERSHIP TRAINING EFFECTS

The elaborated model includes three phases: initial implementation, the intermediate effects on the educational system; and the longer-term effects on either students or staff:

- Implementation begins with the actual training of principals, which results (or not) in principals acquiring new skills.
- Intermediate effects include the resulting changes in the principals' leadership style, which may affect various aspects of the school, including how school staff interact (e.g., increased collaborative planning; clearer sense of purpose among staff) and the overall climate of the school (e.g., a clearer emphasis on academic achievement).
- In the longer term, the changes in staff interactions and in the school climate may affect both school staff and students. Students may respond to the altered school climate with more positive attitudes toward school and/or with increased achievement. Teachers and principals may respond with more positive attitudes about their own profession.

Although at first glance this more elaborated model may appear to complicate the evaluation design, it actually simplifies the task by suggesting specific evaluation questions and indicators of effectiveness. It also underscores the importance of looking at intermediate outcomes that are measurable in the short term. For example, within one year of a training program, one might want to look at the effects of the training on principals' leadership style and on school staff interactions.

Evaluation Questions

The model suggests a set of evaluation questions.

Regarding initial implementation:

(1) What principals attended the training sessions?

(2) What did the principals learn from their participation in the training sessions?

Regarding intermediate effects on the system:

(3) Did what the principals learn affect their leadership styles?

(4) Did changes in principal leadership style affect staff interaction or the climate of the schools in any way?

Regarding longer-term effects:

(5) What were the effects of the changed climate or shifts in teachers' interaction on students' attitudes toward school and on their academic achievement?

(6) Were there any effects on teachers' or principals' sense of professional development?

Indicators of Effectiveness

Given a set of relatively specific evaluation questions, the evaluator can make decisions about the kind of data needed to indicate the effectiveness of the leadership training sessions. Table 3 reviews a set of indicators for each of the evaluation questions as well as a standard of effectiveness for each.

The table includes a wide variety of indicators. Although student achievement outcomes might seem the best indicator of the effectiveness of any Chapter 2-sponsored activity, there are many situations when such data is not appropriate or available. When activities have been in place for a short period, it is not feasible to expect them to result in measurable student-level outcomes. In other cases, the link between the Chapter 2-sponsored activity and student achievement may be very indirect and hard to establish. In these cases, indicators of more immediate effects can tell the evaluator if the project is on track. For example, if principals have changed their leadership style as a result of the leadership training and these changes have improved teachers' attitudes, the evaluator has evidence of positive benefits of the Chapter 2-sponsored activity--even in the absence of student-level data.

Table 3

EVALUATION QUESTIONS AND INDICATORS OF EFFECTIVENESS FOR LEADERSHIP TRAINING PROGRAMS FOR PRINCIPALS

Evaluation Question	Indicator	Standard of Effectiveness
Re: Implementation		
Who attended training?	Number of trainees. Percent of principals attending. Percent of principals targeted.	Reasonable number/percentage given costs and training session goals.
What principals learned?	Principals' knowledge of and attitudes toward material.	Gains in knowledge consistent with the goals of the workshops.
Re: Intermediate Effects		
Effects on principals' leadership style?	How principals involve staff, students, and parents in decisionmaking; principals' manner of communicating goals of school to staff and students.	Change consistent with goals of workshop, district policy and needs, and standards of good practice.
Effects on staff and school climate?	Changes in staff interaction in terms of decisionmaking and planning; changes in teachers' and students' perceptions of central goal of school.	Changes consistent with goals of workshop, district, and standards of good practice; staff and parents' perceptions that decisionmaking process is resulting in effective policies; staff, parents', and students' perceptions that primary goal of school is academic.
Re: Longer-term Effects		
Effects on students' attitudes and achievement?	Teachers' perceptions of shifts in attitudes; student test scores	Improved attitudes toward school; increased focus on academics; higher test scores
Effects on teachers and principals?	Teachers' and principals' sense of worth as professionals; their commitment to their jobs.	Increased sense of self-worth; increased commitment.

Choosing an Appropriate Design and Data Collection Methods

The most useful and convincing evaluation design collects data on a variety of indicators that, when taken together, create a "mosaic of evidence" regarding the effects of a given set of activities.

Such a design typically uses a mix of data collection methods and targets a sample of districts or schools. It is unnecessary to collect evaluative information from all LEAs.

Once the analytical model has been used to generate a set of evaluation questions and to pinpoint the kind of data that will need to be collected (in terms of indicators of effectiveness), the evaluator can make decisions about data collection methods. Following the example of leadership training for principals, we present appropriate data collection methods matched to kind of data one might want to collect:

- LEA Reporting Form. The U.S. Department of Education is providing a model State Annual Reporting Form. States can request that LEAs file similar reports on their use of Chapter 2 funds. These forms can provide SEA evaluators with full counts of the number of staff trained, materials and computer equipment acquired, etc. Although this information will be used primarily for aggregating data for states' annual reports, it also constitutes a base which can be used for evaluation purposes.

- LEA Survey. Surveys of local Chapter 2 directors can provide a great deal of information at a reasonable cost. In most cases, these surveys should be sent to a representative sample of local districts. Section XI in Part Three of this document provides detailed information on sampling procedures. Section XII discusses survey design.

- School Survey. Because most Chapter 2-funded activities are expected to effect what takes place in schools and classrooms, it is advisable that some data be collected at levels below the Chapter 2 director. The school survey can be sent to principals. The survey should focus on specific activities, particularly the degree to which instruction (or other aspects of school life) have changed as a result of the Chapter 2-sponsored intervention.

Table 4

OVERVIEW OF DATA COLLECTION MECHANISMS FOR EVALUATING THE EFFECTS OF LEADERSHIP TRAINING WORKSHOPS

Data Collection Mechanism	Type of Data to be Collected	Target/Respondent
LEA Reporting Form	Figures on dollars spent; number of trainees.	Chapter 2 Director in all LEAs.
LEA Survey	Goals and purposes of training sessions; data on post-workshop evaluations.	Chapter 2 Director in representative sample of LEAs.
School Survey	Principals' perceptions of shifts in leadership style, of changes in staff planning and decisionmaking; principals' opinion of effects on their professional development.	Principals in sample of schools in LEAs that took part in LEA Survey.
Teacher Survey	Teachers' opinions of changes in leadership style, decisionmaking processes and school climate; teachers' opinions of effects of changes on students and teachers.	Sample of teachers in schools participating in School Survey.
Case Studies	Descriptive and explanatory data on all aspects of the implementation, intermediate and (possibly) longer-term effects of leadership training.	Administrators and teachers in small number of districts and schools chosen purposively.
Record Data Review	Student achievement.	Chapter 2 Director as part of LEA Survey.

- Teacher Survey. Ideally, an evaluation of any Chapter 2-funded activity that is meant to affect teaching and learning would measure the effect of the activity on what takes place in the classroom. Surveys of teachers are relatively direct ways to get at this information. Moreover, teachers are the best judges of whether the activity has had any effect on students. Finally, teachers can be asked about the effects of the activity on their own professional development (see Section XII).

- Case studies. Case studies refer to relatively in-depth data collection efforts in a still smaller sample of sites. Typically, they involve on-site data collection that minimally involves interviews) and reviews of record data. In some cases they may include observation of Chapter 2-sponsored activities in action. Case studies are generally labor intensive, but they can result in rich and meaningful descriptions of both the implementation and the effects of a particular set of activities. Moreover, they often provide the kind of information evaluators need to understand why an effort worked or failed.

- Record Data Review. In addition, it is possible to request outcome data (e.g., student achievement) from a set of LEAs. Generally, this should only be done in sites where you are collecting other information. For example, it could be part of the school survey or an integral part of a case study.

The use of any of these data collection methods involves a trade-off between breadth and depth. Collecting data on all material acquisition projects in the state through the LEA Reporting Form provides broad descriptive information on the full population of districts. At the other extreme, carrying out case studies of the use of materials in schools offer rich explanatory data on a small number of sites.

Depending on financial and personnel resources, time, and the nature of the questions the evaluation hopes to answer, evaluators have to make decisions about which data collection methods to employ. Some cases may call for the use of a single evaluation method to collect all the necessary data on a certain type of activity. Here are two examples:

- In State A, three LEAs independently launched leadership training programs for principals. Because the programs seem especially innovative and are clearly linked to the research on effective schooling, the SEA chooses to visit all three sites and carry out in-depth case studies.

- In State B, a number of districts use a portion of their funds to acquire trade books to supplement their reading curriculum. Recognizing that it is very difficult to evaluate the effect of new trade books on any meaningful student or instructional variable, the SEA chooses to collect descriptive data only through the LEA Record Form for these activities.

In most cases, however, it makes sense to strike a balance between depth and breadth by using a range of mechanisms, all aimed at collecting information about similar activities. By using a range of methods, the evaluator is able to collect data on the implementation, intermediate effects, and longer-term effects. The evaluation will be able to include both whole population figures for certain activities (e.g., the number of teachers statewide who participated in staff development efforts) and some information on the effects of those efforts on schools and classrooms in a sample of sites.

Adopting an Approach to Analyzing and Reporting Information

The analysis and reporting of evaluative information should focus on:

- **The implementation of Chapter 2-sponsored activities;**
- **The intermediate effects of the activities on the school system;**
- **The effects of the activities on students and teachers;**
- **The reasons for activities' success or failure (where possible).**

Decisions about data analysis and reporting depend on the type of data that has been collected and the kind of statements evaluators would like to make in their evaluation report. Below, we outline a basic analysis plan. At the end of this section, we include an outline of a state evaluation report.

Compile Basic Descriptive Data on Implementation. Although the central purpose of the evaluation is to describe the effects of Chapter 2-funded activities, any discussion of effects must take place within the context of a

basic description of the activities: for example, how many leadership training sessions were held throughout the state, what were their goals, who participated, and what did they cost. Some of these data may already be aggregated as part of SEAs' annual reports.

Compile Descriptive Data on Intermediate Effects. Next, the analysis should involve a compilation of data (most likely from the the School and Teacher Surveys) concerning the extent to which the leadership training sessions have made any difference in the instruction-related activities of schools. Again, the basic analysis simply involves the summation of numbers and creation of averages. For example, the evaluator might want to report that, *"half of the surveyed teachers reported changes in the overall climate of the school as a result of recent policy changes initiated by their principals. Teachers in elementary schools were twice as likely as teachers in high schools to report such changes."*

This step of the analysis can be bolstered by the use of specific examples from case studies to illustrate major points. For example, one might want to describe a few schools that were visited to give some depth and meaning to the numbers.

Compile Data on the Effects on School Staff and Students. Ultimately, the evaluation has to raise the question of whether the Chapter 2-funded activity had any effect on students or teachers. In some cases (e.g., where a Chapter 2-funded activity has only recently been implemented), an evaluator should not expect to find student-level effects. In other cases (e.g., a program designed to provide direct services to students such as a dropout prevention counseling program), evaluators should expect student-level results fairly soon. Given the data collection methods we discussed above, data on these effects would come from one of three sources:

(1) Teacher or School Survey. These data could be easily compiled through the same mechanisms used in the previous sections. The evaluator here simply reports descriptive results, *"80% of the teachers surveyed reported that they felt more committed to their job as a result of changes in school decisionmaking processes recently implemented by the principal. Only 10% of the surveyed*

teachers reported any effects on students as a result of these changes." Again, one might want to compile data for different subcategories of projects.

(2) Achievement Test Results. Along with any of the survey instruments, the state may ask for supporting documentation in the form of records of student achievement. Typically, we suggest that you treat any such data with extreme caution. The fact that a certain group of students' test scores rose during the same period a computer was introduced into their classroom, for example, does not establish a cause and effect relationship. We suggest you only report this kind of data as part of a case study, during which you will have an opportunity to examine the information in its proper context.

(3) Case Study Data. The case studies should involve interviews with teachers and principals during which you can probe for both opinions and evidence concerning the effects of the introduction of computers. During these visits, evaluators can also review record data, including test results.

Bringing Together Evidence which Explains the Results. A final step in a full evaluation involves attempting to explain why a Chapter 2-sponsored intervention had the effect it did. Again, some data from the surveys may be able to be compiled to help address this issue (*"20% of the surveyed teachers reported that they had no time to participate in the new decision-making teams set up by the principal."*). Full analyses will have to depend upon descriptions of what was found through case study interviews. The small number of case studies do not allow you to make similar statements of incidence, but the evaluator can use the rich data offered through the case studies as instances of broader patterns. For example,

> *"In one district we spoke to teachers in a number of schools who felt their roles had not changed substantially as a result of the principal's participation in a leadership training program. The teachers noted that the principal had returned from the session and made a number of structural changes (e.g., establishing a parent/teacher advisory board grade-level decisionmaking teams). But the teachers were offered no extra time, money, or training to participate in these and none ever got off the ground. These findings suggest that in the absence of sufficient support, teachers cannot be expected to adopt new responsibilities at the school site."*

These analysis steps would allow an evaluator to make meaningful statements about the implementation of a certain set of activities, about the intermediate effects of those materials on the school and its instructional

program, and about the student and teacher-level effects, in certain sites and with certain caveats. In particular, we ask that you pay attention to the following:

Making Statements About "Averages." Because most states have numerous small districts with few students and a few large districts with many students, statements of incidence often have to be stated in more than one manner. For example if in State C, almost all the smaller rural districts are using Chapter 2 funds to purchase computers, while the urban districts are directing funds for other purposes, the statement should read, *"80% of the districts use funds to purchase computers, but these districts only include 20% of the schools and 10% of the students in the state."* Similarly, average district expenditures for computers might be misleading unless broken down by type of district.

For this reason, we suggest that LEA Report Forms and LEA Surveys be coded and keypunched into a computer data base. Many states would be able to handle their data files on a desktop computer with the *Lotus 1-2-3* or *Excel* software. These packages allow easy and quick compilation of data along various dimensions.

Making General Statements Based on Data from a Sample. A second issue regarding the compilation of data concerns estimating whole population parameters with data from sampled districts. That is, is it all right to say that 80% of the districts in the state use computers for mathematics instruction if 80% of the sample of LEAs which received the Survey instrument responded as such. Because of the complexities involved in making such estimates, we suggest that you minimally report: (1) how the survey sample was chosen, and (2) what the percentage of surveyed districts who responded reported. (In Section XI of this document, we provide more detailed advice on estimating population figures from sample data.)

Relating Test Score Changes to a Chapter 2-Sponsored Intervention. We suggest that the assessment of whether a specific intervention actually caused higher test scores only be made when comparison with an appropriate

control group is possible. For example, if Mrs. Jones' 8th grade class gained a significant amount in the CTBS/U mathematics test, these results should be compared with the results of other students who are as similar as possible, except for having been exposed to the Chapter 2-sponsored intervention. Appropriate controls might include the average gain on the same test for Mrs. Jones' previous five classes; the scores of other 8th grade students in the school or the other schools with similar demographics in the district.

Gathering Sufficient Resources to Carry Out the Evaluation

Because the evaluation of the effectiveness of the Chapter 2 program is mandated by Congress, SEAs should use a portion of their state administrative funds to carry out the evaluation.

In addition, there are other ways to tap other state-level evaluation resources.

Carrying out an evaluation of the Chapter 2 program is going to require significant resources and it is imperative that evaluators clarify up front what resources the SEA is going to make available for this effort. Good planning requires concrete information on the amount of funds and manpower that you will have to employ. We cannot suggest an appropriate level of effort for the evaluation, but we do stress that it is not necessary to evaluate each targeted area or every separate activity in the state. Moreover, we offer the following suggestions:

- Make maximum use of existing data collection mechanisms. Most states have a number of mechanisms in place for the communication of information from the local level to the state about the Chapter 2 program or other federal programs. These include the local application for funds, the annual reporting forms, monitoring visits from SEA personnel, and annual meetings. All of these mechanisms can be used to collect data relevant to the design or implementation of the evaluation.

- Use the same methods to evaluate more than one type of activity. The same data collection mechanism can be used to evaluate more than one set or type of activity. For example, a teacher survey could be used to evaluate the effects of a school improvement effort and the use (or non use) of a new set of computers in the school. Similarly, on-site case studies can be used to look at a variety of local projects.

- Sample projects. As we have mentioned, there is no reason that a high quality assessment has to gather evaluative information on the entire population of LEAs in the state. Given limited resources, an evaluation can be much more effective if it focuses on a sample of sites which accurately represent the range of activities across the state. In this case, "less is more." We discuss sampling procedures in Section XI.

- Join forces with other evaluation efforts in the state. Finally, it may be feasible to combine efforts with other state-level evaluation efforts--e.g., travel cost, survey mail cost, just have two parts to everything.

SAMPLE OUTLINE OF STATE EVALUATION REPORT

I. Introduction

A. Overall goals of Chapter 2 program in the state

B. Amount of funds received and allocation procedures

C. Evaluation plan (including rationale for selection of activities for evaluation)

II. Training and Professional Development*

A. State-administered activities
   Funds expended
   Number and types of staff, schools involved
   Evidence of successful implementation

B. Locally administered activities
   Funds expended
   Number and types of staff, schools involved
   Evidence of successful implementation

C. Changes in educational service
   Magnitude, extent, and key patterns of change
   Illustrative examples of change

D. Student- and teacher-level effects
   Magnitude, extent, and key patterns of change
   Illustrative examples of change

E. Reasons for success or failure

III. The Acquisition and Use of Instructional Materials and Computer Equipment*

A. State-administered activities
   Funds expended
   Number and types of students, staff, schools involved
   Evidence of successful implementation

B. Locally administered activities
   Funds expended
   Number and types of students, staff, schools involved
   Evidence of successful implementation

C. Changes in educational service
   Magnitude, extent, and key patterns of change
   Illustrative examples of change

---

* Only include these sections if they were the focus of the state evaluation.

D. Student- and teacher-level effects
   Magnitude, extent, and key patterns of change
   Illustrative examples of change

E. Reasons for success or failure

IV. Direct Services to Students
(including programs designed to enhance the personal excellence of students and student achievement)*

A. State-administered activities
   Funds expended
   Number and types of students, staff, schools involved
   Evidence of successful implementation

B. Locally administered activities
   Funds expended
   Number and types of students, staff, schools involved
   Evidence of successful implementation

C. Changes in educational service
   Magnitude, extent, and key patterns of change
   Illustrative examples of change

D. Student- and teacher-level effects
   Magnitude, extent, and key patterns of change
   Illustrative examples of change

E. Reasons for success or failure

V. Schoolwide Improvement Activities
(including effective schools programs)*

A. State-administered activities
   Funds expended
   Number and types of students, staff, schools involved
   Evidence of successful implementation

B. Locally administered activities
   Funds expended
   Number and types of students, staff, schools involved
   Evidence of successful implementation

C. Changes in educational service
   Magnitude, extent, and key patterns of change
   Illustrative examples of change

---

* Only include these sections if they were the focus of the state evaluation.

D. Student- and teacher-level effects
   Magnitude, extent, and key patterns of change
   Illustrative examples of change

E. Reasons for success or failure

VI. Activities Designed to Meet the Needs of Students at Risk of Failure*

A. State-administered activities
   Funds expended
   Number and types of students, staff, schools involved
   Evidence of successful implementation

B. Locally administered activities
   Funds expended
   Number and types of students, staff, schools involved
   Evidence of successful implementation

C. Changes in educational service
   Magnitude, extent, and key patterns of change
   Illustrative examples of change

D. Student- and teacher-level effects
   Magnitude, extent, and key patterns of change
   Illustrative examples of change

E. Reasons for success or failure

VII. Other Activities Designed to Enhance the Educational Program of the School*

A. State-administered activities
   Funds expended
   Number and types of students, staff, schools involved
   Evidence of successful implementation

B. Locally administered activities
   Funds expended
   Number and types of students, staff, schools involved
   Evidence of successful implementation

C. Changes in educational service
   Magnitude, extent, and key patterns of change
   Illustrative examples of change

---

* Only include these sections if they were the focus of the state evaluation.

SAMPLE OUTLINE OF STATE EVALUATION REPORT (Concluded)

D. Student- and teacher-level effects
   Magnitude, extent, and key patterns of change
   Illustrative examples of change

E. Reasons for success or failure

VIII. Chapter 2-Sponsored Activities in Private Schools

A. The acquisition and use of instructional materials*
   Funds expended, participants
   Changes in educational service
   Student- and teacher-level effects
   Reasons for success and failure

B. Training and professional development*
   Funds expended, participants
   Changes in educational service
   Student- and teacher-level effects
   Reasons for success and failure

C. Direct services to students*
   Funds expended, participants
   Changes in educational service
   Student- and teacher-level effects
   Reasons for success and failure

D. Schoolwide improvement activities*
   Funds expended, participants
   Changes in educational service
   Student- and teacher-level effects
   Reasons for success and failure

E. Activities designed to meet the needs of students at risk of failure*
   Funds expended, participants
   Changes in educational service
   Student- and teacher-level effects
   Reasons for success and failure

IX. Overall Conclusions

Appendices

Methodology
Copies of Instruments
Technical Information

---

* Only include these sections if they were the focus of the state evaluation.

PART TWO: APPROACH TO EVALUATING EACH TYPE OF CHAPTER 2-SPONSORED ACTIVITY

In this part of the Guidance, we discuss the design of evaluations for the different types of Chapter 2-sponsored activities. Following the typology we presented in Table 2, we discuss activities that involve investments in professional development, the acquisition and development of materials and computer equipment, direct services to students, and schoolwide improvement efforts (including effective schools programs) (Sections IV through VII). Following the general outline of our discussion in Section III, for each type of project, we:

- Define the activities that fall within the category.
- Develop an analytical model of how each activity is expected to affect the educational system.
- Discuss design and data collection considerations.
- Review data analysis and reporting issues.

Our discussion here is much less in-depth than that presented in Section III. We refer the reader to that section for more detailed descriptions and analyses.

This part of the Guidance continues with a discussion of the special considerations involved in evaluating activities targeted on at-risk youth (Section VIII), activities for private school students (Section IX), and demonstration projects (Section IX).

IV EVALUATING THE EFFECTIVENESS OF CHAPTER 2 SUPPORT FOR PROFESSIONAL DEVELOPMENT

In this section, we offer suggestions for evaluating the effectiveness of professional development activities. Following the structure we used in Section III, we address the following:

- Defining and selecting activities that fall into this area.
- Specifying an analytical model (including evaluation questions and indicators of effectiveness).
- Choosing appropriate design and data collection methods.
- Adopting an approach to analyzing and reporting information.

Defining and Targeting Professional Development Activities for Evaluation

Traditionally, the bulk of Chapter 2-sponsored professional development activities have involved the inservice training of teachers. These efforts may also include training for district administrators, principals, and instructional assistants. Moreover, for evaluation purposes, state technical assistance to schools and districts can be considered a form of professional development. A reasonable typology of Chapter 2-sponsored professional development activities would include:

Table 5

TYPOLOGY OF PROFESSIONAL DEVELOPMENT ACTIVITIES

	Part of Schoolwide Improvement Efforts? Yes		No	
	Targeted on At-Risk Students? Yes	No	Targeted on At-Risk Students? Yes	No
Technical Assistance to LEAs/Schools				
Teacher Training				
Principal Training				
Other				

Given this typology, evaluators can make decisions about specific evaluation targets. Clearly, the evaluation should include activities sponsored by both the state and local districts. Beyond this, we suggest the following criteria:

- The proportion of funds allocated to the activity.
- The proportion of LEAs conducting the activity.
- Th xtent to which the activity addresses a significant need.

Specifying an Analytical Model for Evaluating Professional Development Activities

Regardless of the specific target of professional development activities (e.g., teachers or principals), the logic of how they affect the education system is similar. In Part One we developed a relatively complex model that outlined how a state-sponsored leadership training program for principals was expected to affect the educational system. Here we provide a similar model regarding teachers (Figure 4):

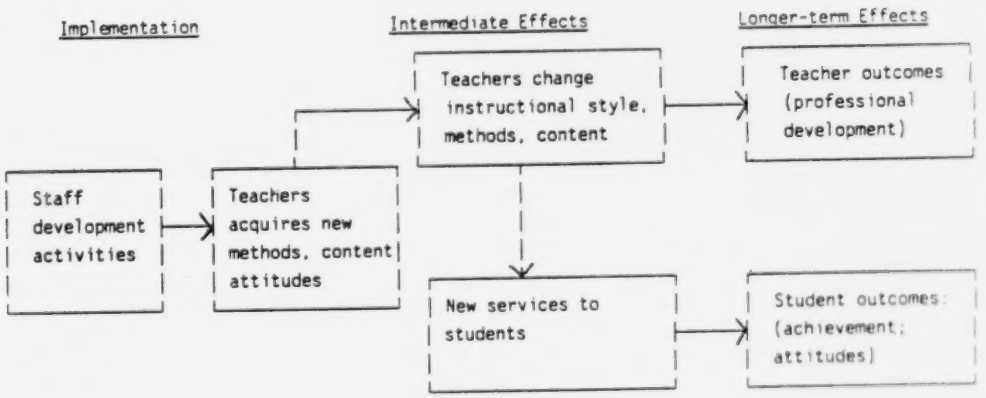


Figure 4

AN ANALYTICAL MODEL OF THE EFFECTS OF PROFESSIONAL DEVELOPMENT ACTIVITIES

The model includes three phases:

- Implementation: the staff development is carried out and teachers do (or do not) acquire more skills, methods, etc.
- Intermediate effects: the staff development results in shifts in how teachers teach and so students' learning experiences change.
- Longer-term effects: changes in students' learning experience result in student-level outcomes (in terms of achievement and attitudes). Changes in teachers' methods and experience may also result in shifts in teachers' own sense of professional development.

This model then describes the chain of events that potentially could result from a high-quality staff training program. It is not meant to imply that evaluations of staff development programs need to collect data on each point in the chain or that these evaluations have to measure student-level outcomes. In fact, the model is meant to underscore the fact that there are meaningful intermediate effects of Chapter 2 about which evaluative data can be collected.

Evaluation Questions

The model is also useful in suggesting a set of reasonable evaluation questions. These questions are analogous to those we developed concerning leadership training in Section III.

Regarding initial implementation:

(1) Who participated in the Chapter 2-sponsored workshops?

(2) Did teachers who participated in the staff development activities acquire new knowledge, attitudes, and/or teaching techniques?

Regarding intermediate effects on the system:

(3) Did teachers' participation in the activities lead them to change their classroom instruction in any way? Any indications that their new methods are superior to the old?

(4) Did students receive appreciably different or better services?

Regarding long-term effects:

(5) What were the effects of the different teaching methods, course content, etc., on students' attitudes toward the subject, enrollment in courses, the amount of time they spend on the subject, and their academic achievement in the subject area?

(6) Were there any effects on teachers' sense of professional development (commitment to teaching, sense of self-worth as teacher, etc.)?

Indicators of Effectiveness

These questions, then, can assist the evaluator to pinpoint the kind of data needed to indicate the effectiveness of the staff development program. These indicators similar to those used to judge the effectiveness of all Chapter 2-sponsored activities. One difference lies in the ability to conduct evaluations immediately following staff development activities and at specified follow-up times with all workshop participants.

Table 6

EVALUATION QUESTIONS AND INDICATORS OF EFFECTIVENESS
FOR PROFESSIONAL DEVELOPMENT ACTIVITIES

Evaluation Question	Indicator	Standard of Effectiveness
Re: Implementation		
Who participated in the staff development effort?	Number and type of teachers	Participating teachers are appropriate given goals of activities and expenditures.
Did teachers learn what was presented?	Number/percent of teachers who evidenced having learned.	Reasonable number/percent, given goals and level of expenditures.
Re: Intermediate Effects		
Have teachers altered their teaching? So students receive better services?	Change in what/how teachers teach and in what students experience.	Change consistent with goals of training, with district policy and goals, and with notions of good practice; teachers' opinion that change is effective.
Re: Longer-term Effects		
Student outcomes (attitudes, achievement)?	Student test scores; teachers' perceptions of changes in students' performance.	Higher standardized test scores; teacher ratings that student performance has improved; teacher ratings that student attitudes toward subject have improved.
Teacher outcomes?	Teachers' sense of professional worth and commitment to their job.	Increased professional activities; increased interest in profession; increased sense of self-worth as professional.

Choosing Design and Data Collection Methods

Table 7 provides an overview of the type of data collection mechanisms that can be used to assess professional development activities. It also includes the respondent or target of each mechanism and the type of data it can be used to gather.

Table 7

OVERVIEW OF DATA COLLECTION MECHANISMS FOR PROFESSIONAL DEVELOPMENT ACTIVITIES

Data Collection Mechanism	Type of Data to be Collected	Target/Respondent
LEA Reporting Form	Figures on dollars spent; number of teachers participating.	Chapter 2 Director in all LEAs.
Post-training Session Evaluation Form	Teachers' perceptions of the value of staff development activities; change in teachers' knowledge, attitudes, or ideas about teaching.	Teachers participating in training session.
LEA Survey	Goals and purposes of training; more detailed data on the effects of the training.	Chapter 2 Director in representative sample of LEAs.
School Survey	Data on effects at the school site; degree to which instructional practice has changed.	Principal in sample of schools in LEAs that took part in LEA Survey.
Teacher Survey	Detail on changes in instruction and on effects on both students and teachers.	Sample of teachers in schools participating in School Survey.
Case Studies	Descriptive and explanatory data on all aspects of the effects of the training on classrooms, teachers, and students.	Administrators and teachers in small number of districts and schools chosen purposively.
Record Data Review	Student achievement.	Chapter 2 Director as part of LEA Survey.

Adopting an Approach to Data Analysis and Reporting

The analysis and reporting of data regarding professional development activities should follow the general outline we reviewed in the previous section (III). That is, the analysis and reporting of evaluative information should focus on:

- The implementation of Chapter 2-sponsored activities (e.g., how many teachers were trained, for how much money, in what subject areas, etc.).
- The intermediate effects of the activities on the school system (e.g., data on changes in teachers' practice and students' experience).
- The effects of the activities on students and teachers (in terms of achievement and attitudes).
- The reasons for activities' success or failure (where possible).

More detailed discussion of analysis and reporting functions can be found in Section III.

V EVALUATING THE EFFECTIVENESS OF THE ACQUISITION AND DEVELOPMENT OF MATERIALS AND COMPUTER EQUIPMENT

In this section, we offer suggestions for evaluating the effectiveness of activities involving the acquisition or development of material and computer equipment. Following the structure we established in the previous section (III), we address the following:

- Defining and selecting activities that fall into this area.
- Specifying an analytical model (including evaluation questions and indicators of effectiveness).
- Choosing appropriate design and data collection methods.
- Adopting an approach to analyzing and reporting information.

Defining and Selecting Material and Computer Equipment Acquisition and Development Activities for Evaluation

The acquisition of materials and computer equipment, especially library books and computers, has been a common use of Chapter 2 funds at the local level. Typical funded activities have included:

- The purchase of computer hardware.
- The purchase or development of computer software.
- The purchase or development of other curricular materials.
- The purchase of library books
- The purchase or development of other materials.

In many instances, the acquisition or development of materials and computer equipment has been part of a broader school improvement program and/or of efforts to improve educational services for at-risk students. An efficient way to categorize these activities involves modifying the matrix we presented in Part One:

Table 8

TYPOLOGY OF MATERIAL AND COMPUTER EQUIPMENT ACQUISITION AND DEVELOPMENT ACTIVITIES

	Part of Schoolwide Improvement Efforts? Yes		Part of Schoolwide Improvement Efforts? No	
	Targeted on At-Risk Students? Yes	Targeted on At-Risk Students? No	Targeted on At-Risk Students? Yes	Targeted on At-Risk Students? No
Computer hardware				
Computer software				
Curricular materials				
Library books				
Other				

Given this typology of material acquisition and development activities, evaluators can make decisions about where to invest scarce evaluation resources. Again, the following criteria may be the most useful:

- The proportion of funds allocated to the activity.
- The proportion of LEAs conducting the activity.
- The extent to which the activity addresses a significant need.

Specifying an Analytical Model for Evaluating the Acquisition or Development of Material and Computer Equipment

Regardless of the specific type of material or computer equipment, the logic of the effect of their acquisition or development on the educational system is similar. We present an analytical model that describes this process below. We then indicate how it can be used to devise evaluation questions and to pinpoint indicators of effectiveness.

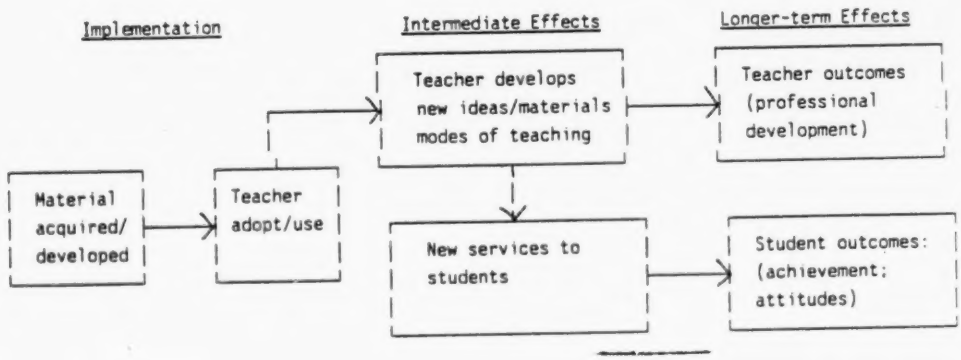


Figure 5

AN ANALYTICAL MODEL OF THE EFFECTS OF MATERIAL ACQUISITION AND DEVELOPMENT

The model includes three phases:

- Implementation: the acquisition of new material or computer equipment, which are adopted (or not) by teachers, principals, or other staff.
- Intermediate effects: the new material or computer equipment then may lead teachers to change what or how they teach, changing (or not) students' learning experience.
- Longer-term effects: the students' experience with the new materials or computer equipment, in turn, may lead to student level outcomes, such as changed attitudes, increased achievement, or decisions about persistence in a subject or field. Teachers' experience with the new material may also affect their feelings about their job, causing them to alter their methods of teaching, or prompting them to learn more about a particular issue.

Evaluation Questions

As we discussed in Part One, the development of an analytical model can assist the evaluator to devise specific evaluation questions that are both meaningful and answerable. Here are a set of evaluation questions suggested by this model:

Regarding initial implementation:

(1) What material or computer equipment was acquired? In what kind of LEAs?

(2) What number and percentage of the target teacher population adopted the new material (and why)?

Regarding intermediate effects on the system:

(3) Did adoption of the material change how or what the adopting teachers teach? If so, what changed?

(4) Did students receive appreciably different (or better) services?

Regarding longer-term effects:

(5) What were the effects of the services or the use of different material on students' attitudes toward the subject, on the amount of time they spend on the subject, on course enrollments, and on their academic achievement in the subject area.

(6) Were there any effects on teachers' sense of professional development?

Indicators of Effectiveness

Given these questions, the evaluator can make decisions about the kind of data needed to indicate the effectiveness of the investment in particular materials or computer equipment. Table 9 points to a set of possible indicators.

Table 9

EVALUATION QUESTIONS AND INDICATORS OF EFFECTIVENESS
FOR THE ACQUISITION AND DEVELOPMENT OF MATERIAL AND COMPUTER EQUIPMENT

Evaluation Question	Indicator	Standard of Effectiveness
Re: Initial Implementation		
What material was acquired or developed?	List and examples of material and computer equipment.	Acquisition bears some relation to reasonable LEA/school goals.
Who adopted material or computer equipment?	Number/percent of teachers who adopted.	Reasonable number/percent, given goals and level of expenditures.
Re: Intermediate Effects		
Have teachers altered their teaching? Do students receive better services?	Change in what/how teachers teach and in what students experience.	Change consistent with goals of project, with district policy and needs, and with notions of good practice; teachers' opinion that change is effective.
Re: Longer-term Effects		
Student outcomes (attitudes, achievement)?	Student test scores; teachers' perceptions of changes in students' performance.	Higher standardized test scores; teacher ratings that student performance has improved; teacher ratings that student attitudes toward subject have improved.
Teacher outcomes?	Teachers' sense of professional worth and commitment to their job.	Increased professional activities; increased interest in profession; increased sense of self-worth as professional.

Choosing Design and Data Collection Methods

Table 10 provides a brief outline of appropriate data collection mechanisms, their targets (or respondents), and the type of information they can be used to gather.

Table 10

OVERVIEW OF DATA COLLECTION MECHANISMS FOR MATERIAL AND COMPUTER EQUIPMENT ACQUISITION AND DEVELOPMENT

Data Collection Mechanism	Type of Data to be Collected	Target/Respondent
LEA Reporting Form	Figures on dollars spent; computer equipment and material acquired; number of teachers and and students using it.	Chapter 2 Director in all LEAs.
LEA Survey	Goals and purposes of material acquisition; more detailed data on adoption and use.	Chapter 2 Director in representative sample of LEAs.
School Survey	Data on adoption and use at school site; degree to which instructional practice has changed.	Principal in sample of schools in LEAs that took part in LEA survey.
Teacher Survey	More detail on adoption; report on changes in instruction and on effects on both students and teachers.	Sample of teachers in schools participating in School Survey.
Case Studies	Descriptive and explanatory data on all aspects of material acquisition, adoption, use in classrooms and effects of teachers and students.	Administrators and teachers in small number of districts and schools chosen purposively.
Record Data Review	Student achievement.	Chapter 2 Director as part of LEA Survey.

Adopting an Approach to Data Analysis and Reporting

The analysis and reporting of data regarding materials and computer equipment acquisition and development should follow the general outline we reviewed in Section III of this document. That is, the analysis and reporting of evaluative information should focus on:

- The implementation of Chapter 2-sponsored activities (e.g., how many computers were purchased and how much they cost; who is using the hardware, for what purpose, at what grade levels, in what subject areas, etc.).
- The intermediate effects of the activities on the school system (e.g., data on the extent to which the use of computers has made any difference in the instructional activities of the school).
- The effects of the activities on students and teachers (in terms of achievement and attitudes).
- The reasons for activities' success or failure (where possible).

Note that analysis of data on the acquisition or development of material should include a description of the use of the material within the context of an activity aimed at one of the targeted areas. In the case of the purchase of computer hardware, however, the analysis should address the use of the computers within the context of the instructional program.

VI EVALUATING THE EFFECTIVENESS OF DIRECT SERVICES TO STUDENTS

In this section, we offer suggestions for evaluating the effectiveness of activities that provide services directly to students. Following the previous sections, we address the following:

- Defining and selecting activities that fall into this area.
- Specifying an analytical model (including evaluation questions and indicators of effectiveness).
- Choosing appropriate design and data collection methods.
- Adopting an approach to analyzing and reporting information.

Defining and Selecting Activities that Provide Direct Services to Students for Evaluation

There are numerous efforts funded by Chapter 2 dollars that provide services directly to students. Typically, these activities fall within the targeted purpose entitled "Programs to enhance personal excellence and student achievement." These would include direct instruction to students in ethics, the performing and creative arts, humanities, health education, and physical fitness. This broad category also includes some activities that fall under the heading "Programs to enhance school climate and educational programs." These include direct instruction for gifted and talented students, early childhood education, and various counseling services. A reasonable typology of these varied activities would include categories for the type of service (e.g., instruction vs. counseling) and the target of the services (e.g., at-risk students vs. gifted and talented students). Here, then, we present a matrix somewhat different from the one presented in previous sections of the guidance:

Table 11

TYPOLOGY OF ACTIVITIES THAT PROVIDE DIRECT SERVICES TO STUDENTS

	Target of Activity			
	At-Risk Students	Gifted and Talented	Preschool Students	Other
Instruction				
Counseling				
Other				

As in the case of the previous typologies we discussed, it may be important to maintain records on other program dimensions such as subject area and grade level. Again, selection of activities to evaluate should turn on some reasonable criteria, such as the proportion of funds or districts involved or the importance of the activity to state policymakers.

Specifying an Analytical Model for Evaluating Activities that Provide Direct Services to Students

In the case of direct services to students, the analytical model is very relatively straightforward, reflecting the proximity of the Chapter 2-funded effort to the intended beneficiary, the student:

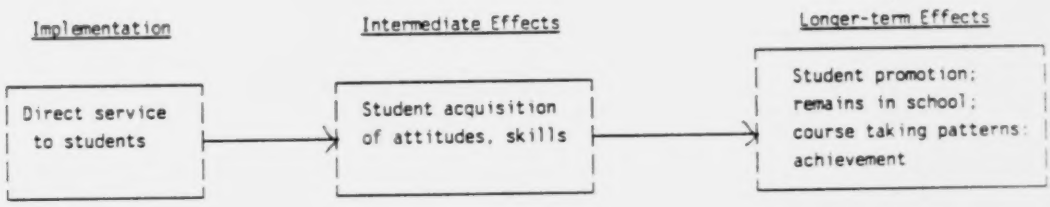


Figure 6

AN ANALYTICAL MODEL OF THE EFFECTS OF ACTIVITIES THAT PROVIDE DIRECT SERVICES TO STUDENTS

This model, like the previous ones we have outlined, includes three phases:

- Implementation: services are provided to students (or not).
- Intermediate effects: the student takes from the service some benefit the service was intended to bestow (e.g., new skills or attitudes).
- Longer-term effects: the student changes some longer-term behavior, such as deciding to remain in school or taking a certain type of course.

For example, a group of districts might provide Saturday morning instruction to gifted students in high schools. In terms of implementation, the evaluator is interested in whether the classes ever take place and if students attend. In terms of intermediate effects, the evaluator is interested in changes in attitudes and ability (e.g., did the student learn about physics?). In terms of longer-term effects, the evaluator might be interested in whether the students changed their choice of courses in their remaining semesters, or shifted potential college majors.

Evaluation Questions

Appropriate evaluation questions are:

Regarding initial implementation:

(1) What direct services were provided?

(2) What number and percentage of the targeted student population participated?

Regarding intermediate effects:

(3) What did the students take away from the direct services in terms of skills and attitudes?

Regarding longer-term effects on the students:

(4) What were the effects of the services on students' school-related activities (e.g., course selection; academic achievement; staying in school; promotion).

Indicators of Effectiveness

Given these questions, the evaluator can make decisions about the kind of data needed to indicate the effectiveness of the investment in particular services. Table 12 points to a set of possible indicators.

Table 12

EVALUATION QUESTIONS AND INDICATORS OF EFFECTIVENESS OF ACTIVITIES THAT PROVIDE DIRECT SERVICES TO STUDENTS

Evaluation Question	Indicator	Standard of Effectiveness
What services were provided?	List and examples of services.	Services bears some relation to reasonable LEA/school goals.
Who participated?	Number/percent of targeted students.	Reasonable number/percent, given goals and level of expenditures.
What students acquired from the services?	Student responses in interview; teacher perceptions.	Evidence of having acquired the skills or attitudes the services were meant to engender.
Student outcomes (attitudes, achievement, participation in courses, etc.)	Student test scores; teachers' perceptions of changes in students' performance and attitudes.	Higher standardized test scores; teacher ratings that student performance has improved; teacher ratings that student attitudes toward subject have improved.

Choosing Design and Data Collection Methods

Table 13 provides a brief outline of appropriate data collection mechanisms, their targets (or respondents), and the type of information they can be used to gather.

Table 13

OVERVIEW OF DATA COLLECTION MECHANISMS FOR ACTIVITIES THAT PROVIDE DIRECT SERVICES TO STUDENTS

Data Collection Mechanism	Type of Data to be Collected	Target/Respondent
LEA Reporting Form	Figures on dollars spent; number of participating students.	Chapter 2 Director in all LEAs.
LEA Survey	Goals and purposes of services; some detail on effects.	Chapter 2 Director in representative sample of LEAs.
School Survey	Data on implementation; principals' perceptions of effects.	Principal in sample of schools in LEAs that took part in LEA Survey.
Teacher Survey	Teachers' perceptions of student-level effects.	Sample of teachers in schools participating in School Survey.
Case Studies	Perceptions of effects on students; concrete examples of these effects.	Administrators, teachers, and students in small number of districts and schools chosen purposively.
Record Data Review	Student achievement.	Chapter 2 Director as part of LEA Survey.

Adopting ah Approach to Data Analysis and Reporting

The analysis and reporting of data regarding direct services to students should follow the general outline we have reviewed in the previous three sections (III, IV, and V). That is, the analysis and reporting of evaluative information should focus on:

- The implementation of Chapter 2-sponsored activities (e.g., how many students were involved, how much money was spent, in what subject areas, etc.).
- The effects of the activities on students (in terms of achievement and attitudes).
- The reasons for activities' success or failure (where possible).

VII EVALUATING THE EFFECTIVENESS OF SUPPORT FOR SCHOOLWIDE IMPROVEMENT, INCLUDING EFFECTIVE SCHOOLS PROGRAMS

In this section, we offer suggestions for evaluating the effectiveness of Chapter 2-sponsored school improvement activities (including effective schools programs). Following the previous sections, we address the following:

- Defining and selecting activities that fall into this area.
- Specifying an analytical model (including evaluation questions and indicators of effectiveness).
- Choosing appropriate design and data collection methods.
- Adopting an approach to analyzing and reporting information.

Defining and Selecting School Improvement and Effective Schools Activities

A major thrust of many state and local Chapter 2-sponsored activities involve school improvement and effective schools efforts. Indeed, the current legislation calls for states to use a portion of their funds to support effective schools programs.

"School improvement" is a generic term that refers to any effort to improve the school as a whole, that is, activities that are not targeted on a subset of students, a particular grade level, or subject area. "Effective schools programs" is a specific term that refers to programs that have the objectives of:

(1) Promoting school-level planning, instructional improvement, and staff development.

(2) Increasing the academic achievement of all children and particularly educationally deprived children.

(3) Achieving as ongoing conditions in the school the following factors:

- Strong and effective administrative and instructional leadership.
- Emphasis on the acquisition of basic and higher-order skills.
- A safe and orderly school environment.
- A climate of expectation that virtually all children can learn under the appropriate conditions.
- Continuous assessment of student and programs to evaluate the effects of instruction.

The actual mechanisms by which state or local education agencies use Chapter 2 funds to support school improvement efforts will vary considerably and will typically involve the activities we have discussed in the previous section (staff development, material acquisition, and direct services to students). In operational terms, then, the evaluation of school improvement efforts should follow the outline presented in the previous three sections.

In the case of effective schools programs, we can specify a few additional activities, which we have not addressed in the previous sections. These can be differentiated by whether the program is targeted on schools that serve high concentrations of at-risk students:

Table 14

TYPOLOGY OF EFFECTIVE SCHOOLS PROGRAMS

	Targeted on Schools with High Concentrations of At-Risk Students?	
	Yes	No
Professional development (and technical assistance)		
Acquisition/development of materials and computer equipment		
Development/implementation of evaluation systems		
Improvement of parent and community involvement		

As in the case of the previous typologies we discussed, selection of activities to evaluate should turn on some reasonable criteria, such as the proportion of funds or districts involved or the importance of the activity to state policymakers.

Specifying an Analytical Model for Evaluating School Improvement and Effective Schools Programs

To the extent that school improvement activities consist of staff development efforts, the acquisition or development of materials, or direct services to students that are targeted at the entire school instead of a part of it, we can utilize the same analytical models that are developed in the previous three sections. In the case of effective schools programs, however, an additional model is called for because these programs are so well defined in the legislation:

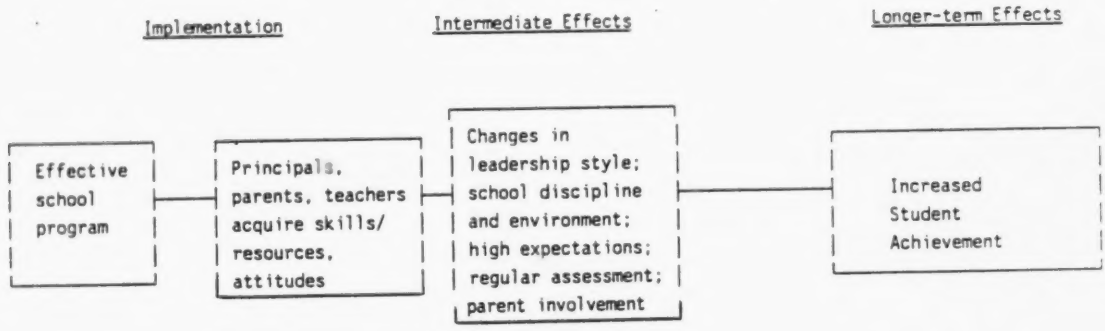

Implementation
Intermediate Effects
Longer-term Effects
Effective school program
Principals, parents, teachers acquire skills/ resources, attitudes
Changes in leadership style; school discipline and environment; high expectations; regular assessment; parent involvement
Increased Student Achievement

Figure 7

AN ANALYTICAL MODEL OF THE EFFECTS OF EFFECTIVE SCHOOLS PROGRAMS

This model, then, involves three separate phases:

- Implementation: in which the effective school effort takes place and school staff develop new skills, resources, and/or attitudes.
- Intermediate Effects: these new attitudes, skills, and resources may result in schoolwide changes in leadership style, academic climate, etc.

- Longer-term Effects: these changes in the school may ultimately result in increased student achievement, the goal of all effective school programs.

Evaluation Questions

Appropriate evaluation questions suggested by this model are:

Regarding initial implementation:

(1) What staff, parents, and schools participated in effective school efforts?

(2) What did the schools' staff learn from or take away from their participation in the effort (in terms of skills, attitudes, and resources)?

Regarding intermediate effects:

(3) What were the effects of the effort on the school as a whole (in terms of leadership, climate, expectations, and evaluation)?

Regarding longer-term effects:

(4) Did effective school programs result in higher student achievement?

Indicators of Effectiveness

These questions suggest specific data to be collected (Table 15):

Table 15

EVALUATION QUESTIONS AND INDICATORS OF EFFECTIVENESS OF EFFECTIVE SCHOOLS PROGRAMS

Evaluation Question	Indicator	Standard of Effectiveness
Re: Implementation		
Who participated in the programs or efforts?	Number and type of staff/parents/teachers.	Participants are appropriate given goals of program.
Did participants acquire desired skills, attitudes, resources?	Number/percent who did so.	Reasonable number/percent, given goals and level of expenditures.
Re: Immediate Effects		
Have appropriate changes taken place in school? different (better) services?	Change in climate, leadership, discipline, expectations, and evaluation procedures.	Change consistent with goals of program, district policy and needs, and the literature on effective schools (e.g., a school climate focused on academic achievement).
Re: Longer-term Effects		
Student outcomes (attitudes, achievement)?	Student test scores; teachers' perceptions of changes in students' achievement.	Higher standardized test scores; teacher ratings that student performance has improved; or other alternative measures.

Choosing Design and Data Collection Methods

Table 16 provides an overview of the type of data collection mechanisms that can be used to assess effective schools programs.

Table 16

OVERVIEW OF DATA COLLECTION MECHANISMS FOR EFFECTIVE SCHOOLS PROGRAMS

Data Collection Mechanism	Type of Data to be Collected	Target/Respondent
LEA Reporting Form	Figures on dollars spent; number of participants.	Chapter 2 Director in all LEAs.
LEA Survey	Goals and purposes of programs; detail on what participants acquired.	Chapter 2 Director in representative sample of LEAs.
School Survey	Data on effects at the school site; degree to which effective school characteristics are present.	Principal in sample of schools in LEAs that took part in LEA Survey.
Teacher Survey	Detail on school-level changes and on effects on teachers.	Sample of teachers in schools participating in School Survey.
Case Studies	Descriptive and explanatory data on all aspects of the effects of the program on classrooms, teachers, and students.	Administrators and teachers in small number of districts and schools chosen purposively.
Record Data Review	Student achievement.	Chapter 2 Director as part of LEA Survey.

Adopting an Approach to Data Analysis and Reporting

The analysis and reporting of data regarding school improvement and effective schools programs should follow the general outline we reviewed in the previous sections (III, IV, and V). That is, the analysis and reporting of evaluative information should focus on:

- The implementation of effective school efforts (how many schools and staff involved; what staff learned, etc.).
- The intermediate effects of the activities on the school (specifically in terms of climate, expectations, leadership, evaluation, and discipline).
- The effects of the activities on students achievement.
- Reasons for success or failure.

VIII EVALUATING THE EFFECTIVENESS OF CHAPTER 2-SPONSORED ACTIVITIES DIRECTED AT AT-RISK STUDENTS

Throughout the Chapter 2 legislation, Congress has confirmed its support for activities targeted on youth at-risk of school failure. State Chapter 2 applications suggest that many SEAs and LEAs are launching efforts to address the needs of this population. If such efforts constitute a large or especially innovative part of the overall Chapter 2 program in a state, it may make sense to focus a part of the SEA evaluation on these activities.

Designing an appropriate evaluation for activities targeted on at-risk students poses no difficulties that we have not already covered in Sections III through VII of this document. At-risk students are only a target of a Chapter 2-sponsored activity; the activity itself--whether it involve staff development (see Section IV), material acquisition (Section V), direct student services (Section VI), or effective schools efforts (Section VII)--can be evaluated through the designs we have reviewed in the previous sections.

State agencies may encounter some problems in deciding what activities are clearly targeted on this population and which are not. We suggest that states follow the legislative criteria, that these be activities that are targeted on students who are at risk of failure in school and of dropping out and for whom providing an education entails higher than average costs. (See 20 U.S.C. 2941.)

IX EVALUATING THE ACHIEVEMENT OF OTHER LEGISLATIVE GOALS UNDER CHAPTER 2

Because the Chapter 2 program provides a variety of services to many different populations, evaluations of its effectiveness have to take into account a number of special considerations. In this section, we discuss the special cases of:

- Services for students in private schools.
- The use of Chapter 2 dollars for "seed money" versus continuing support.

Evaluating the Contribution of Chapter 2 to Private School Students

SEA evaluations must collect data on services to private school students.

The Chapter 2 legislation requires the equitable participation of students enrolled in private schools and state evaluations of the program should address the use of funds for this population. The vast majority of local educational agencies that have eligible private schools have established relationships with these schools and have arranged for ongoing, Chapter 2-funded services. Yet, LEAs face a number of difficulties in evaluating the use of Chapter 2 funds for private school students:

- Funds for private school students often constitute a small fraction of a district's Chapter 2 budget.
- The logistics of collecting evaluation data from private schools are often complex.

At the same time, however, the evaluation tasks involved in assessing the use of funds to serve private school students is no different than that involved when Chapter 2 supports efforts for public school students. That

is, evaluation should use the same basic approach in assessing the acquisition of instructional materials for private school students as they would for public school students--they simply need to collect an additional set of data. The issues with such data collection are primarily operational. Although some descriptive data and perceptions of impact can be solicited from LEA Chapter 2 coordinators, most indicators of effectiveness imply school or teacher surveys, case studies, or other collection of data from the private schools themselves.

Yet, many SEAs and LEAs face distinct limitations on their ability to gather such information at reasonable cost. So, we offer the following advice:

- States may wish to be selective in the amount of effort they put into the evaluation of services for private school students. Direct data collection from private schools should be done when a substantial proportion of funds goes to these services.
- Sampling private schools should be done judiciously. For example, states may wish to eliminate from the evaluation any private school receiving only a minimal amount of Chapter 2 funds.
- Wherever possible, use established hierarchies to reach private schools. For example, in many areas, it will make sense to consult with the Roman Catholic Archdiocese. It may even be possible to send surveys through organizations that oversee some set of private schools.

Evaluating the Effectiveness of Chapter 2-Supported Demonstration Projects

Since its inception, the Chapter 2 program has been designed to provide both the initial funding for new and promising demonstration projects and funds for the continuing supplementary support of innovative activities. Generally, the same evaluation frameworks and methods that we have discussed in Sections IV through VII apply equally to both types of activities. However, evaluations of model programs or demonstration projects for which Chapter 2 has provided initial funding often will call for a different emphasis in the criteria used to judge them. In particular:

- Demonstration projects should not be judged by the number of students, teachers, or schools affected. By their very nature, demonstration projects are meant to be limited in scope. They are not meant to have widespread effects in themselves.

- Demonstration projects should be able to show effects on students and/or teachers. If demonstration projects are to serve as a model for other schools or districts, they have to offer evidence of results. Thus, evaluations of these projects should include detail on educational outcomes.

- Demonstration projects may also be judged on the extent to which they have served as a model for other schools or districts. The ultimate purpose of demonstrations projects is to serve as a model for educators in other settings. Thus, one important criteria of demonstration projects that have been in existence for a few years concerns the extent to which they have been copied elsewhere.

Chapter 2 funds may also serve the more modest goal of supporting the initiation of activities that are subsequently picked up by local funding. This kind of "institutionalization" of Chapter 2-supported activities provides another indicator of the success of Chapter 2 funds as seed money.

For example, a state may choose to use Chapter 2 funds to support a model dropout prevention program in three LEAs. It makes little sense to measure the effects of the demonstration project by the number of students served--which in proportion to the state as a whole, will be very low. At the same time, model projects are supposed to demonstrate the effectiveness of certain approaches to a problem. Consequently, within a year of its implementation, the evaluator should try and gauge the effects of the program on students in the participating schools. Finally, after the project has been in existence for a few years, evaluators should begin to look for evidence of its spread to other LEAs and schools and its institutionalization in the three original districts.

PART THREE: OPERATIONAL SUGGESTIONS AND USEFUL REFERENCES

In this final part of the document, we provide more detailed guidance on specific design issues. We review the issues involved in selecting an appropriate sample, designing a useful survey, and carrying out an effective case study. Finally, we provide a list of resources for those evaluators desiring more in-depth information on various design and analysis matters.

X SAMPLING STRATEGIES

Throughout this document we have underscored the utility of using samples of districts, schools, and teachers to assess the effect of Chapter 2-sponsored activities. Generally, it is neither methodologically necessary nor practical to study the entire population affected by the intervention under investigation. In those instances in which the population is small (e.g., 50 science teachers involved in a special state program) and the unit cost of data collection is low (e.g., a self-administered questionnaire), it may make sense to study the entire population. However, in most cases, the large size of the population and the high cost of data collection make it necessary to sample respondents. In this section we offer some practical advice on how to choose samples.

The two major issues evaluators have to deal with when constructing samples are bias and precision:

> Bias refers to the representativeness of the sample: are the respondents to your survey representative of the relevant population? If half of the teachers you survey say they are using the new computers in their classrooms, does that mean that half the teachers who received computers are using them? It may not if the way you choose the sample assured that you reached more computer users than non-users, or if a greater portion of sampled computer users responded to the survey than non-users. Bias can be minimized (1) by using a list to select the sample that is representative and exhaustive of the population to which you wish to generalize, (2) by randomly* selecting from this list, and (3) by obtaining a high response rate from those selected.
>
> Precision refers to the accuracy of the estimates derived from the sample. Even if you have a totally unbiased sample, because it is just a part of the full population, data from it represents an estimate of what is true in the population as a whole. Precision is a function of the sample size.

---

* Note, random is not the same as arbitrary or haphazard. In a random sample the probabilities of selection are known.

In the following discussion we address both of these issues in relation to samples of various populations (LEAs, schools, etc.). We emphasize common sense steps evaluators can take to minimize bias and maximize precision.

The discussion that follows refers to survey samples. We discuss samples for case studies in Section XII.

Sampling Districts

There are two major questions to be considered in making district sampling decisions: how many districts to include and which districts these should be.

Overall Sample Sizes

The number and size of districts within states varies so considerably that is is difficult to make general comments about the appropriate number of districts to include in a sample. If the number of districts in the state is small, it may make sense to include all districts for survey purposes. Table 17 provides some general guidance on the appropriate number of LEAs to include in general surveys, assuming different populations of districts across states.

Table 17

CHOOSING OVERALL SAMPLE SIZES FOR DISTRICT SURVEYS

Number of Districts In State	Appropriate* Sample Size
50	35
100	50
300	75
500+	100

* Sample size needed to obtain a maximum sampling error of +/- 10% at the 95% confidence interval assuming a simple random sample. This means that in 95 samples out of a 100, the estimated percentage of districts responding in a certain way will be within 10 points of the "true" percentage--what would have been obtained if the entire population had been surveyed.

Sample Designs

Given a target number of districts, the evaluator can choose to utilize a simple random sample or a stratified random sample. In either case, the first step in the sampling process is to develop a list of districts from which to select the sample. This can include all districts in the state or all districts with particular characteristics (e.g., districts that use Chapter 2 for professional development). For the sample to be representative, however, the list will need to include all districts about which you plan to generalize.

Simple Random Samples. The most straightforward sampling method is a simple random one. To select a simple random sample, you assign a number to each district on the list and then use a random number table to select the desired number of districts. Unless you plan to use a computer, it is more practical to select a systematic random sampling--a sample in which districts on the list are selected at fixed intervals (e.g., every 5th district, every 10th district). To ensure against bias, the first district on the list should be selected at random. Also, the order of the list should be examined to make sure it does not bias the selection (e.g., every 5th district on the list is an urban district). It is unlikely to bias the selection if districts are organized in alphabetical order before sampling.

Stratified Random Samples. For policy reasons, simple random samples often do not make sense. The statistical virtue of the simple random sample--that all districts in a state have an equal chance of being selected--is its policy downfall. Especially in states with a few large districts (which receive a large portion of the Chapter 2 funds) and many small districts, simple random samples too often will exclude most if not all the large districts.

Consequently, it often makes sense to sort the population of districts into groups by certain characteristics (e.g., size and poverty level) before sample selection. A higher proportion of districts can then be selected from categories of importance (e.g., large, high-poverty districts). This method

of stratification allows the evaluator to ensure that the sample includes districts of importance. Within these strata, districts should still be chosen randomly.

There are, however, two drawbacks to stratified samples. First, because they include a disproportionate number of districts from certain strata, the sample will have to be larger to obtain the same precision as a simple random sample. Second, as we discuss in more detail at the end of this section, data will need to be weighted for the results to be generalizable to the entire population before analysis.

To categorize districts by enrollment, we typically use the following strata, based on national population figures:

- Very Large (25,000 students or more)
- Large (10,000 to 24,999 students)
- Medium (2,500 to 9,999 students)
- Small (600 to 2,499 students)
- Very small (less than 600 students)

School Samples

The selection of school samples can be more problematic. Many states do not have access to a list or database of all the schools, and the population of schools is much larger than for districts. For a variety of design reasons, however, the choice of schools can be simplified. We suggest the following:

- Target only schools in those districts that participate in the LEA Survey.
- Within those districts, exclude schools that have not participated in any Chapter 2-sponsored activities.
- Stratify the sample of schools within district by grade level served.
- Choose among the remaining schools randomly.

In practical terms you can request that district coordinators carry out the sampling process.

The number of schools to include will vary according to the number of districts, their size etc. We suggest the following guidelines:*

- If possible, randomly choose 100 or more schools overall.
- Ask local coordinators to randomly choose 10% of the district schools, and limit the possible number of schools per district to 20.

Sampling Teachers

We have suggested in a number of places that certain indicators of effectiveness can be best collected through teacher surveys. There are basically two methods of reaching a sample of teachers. First, it may be possible to target a population of teachers, all of whom have participated in a particular workshop or program sponsored by Chapter 2. In this case, it may make most sense to choose a random sample from the list of participants (or if the number is small) to survey them all. In other cases, however, the state will not have access to this type of information. In these instances, we suggest that the teacher sample be nested within schools which are completing school surveys. Under the best circumstances, the SEA would have a list of all teachers in these schools and choose a sample of them. In most cases, however, the state will have to send the Teacher Surveys along with the School Survey and ask the principal to choose teachers along the following decision rules:

- Select only those teachers who have participated in Chapter 2-sponsored activities during the past year.
- Choose teachers at particular grade levels (depending on the target of the evaluation).
- Choose no more than one teacher per grade.

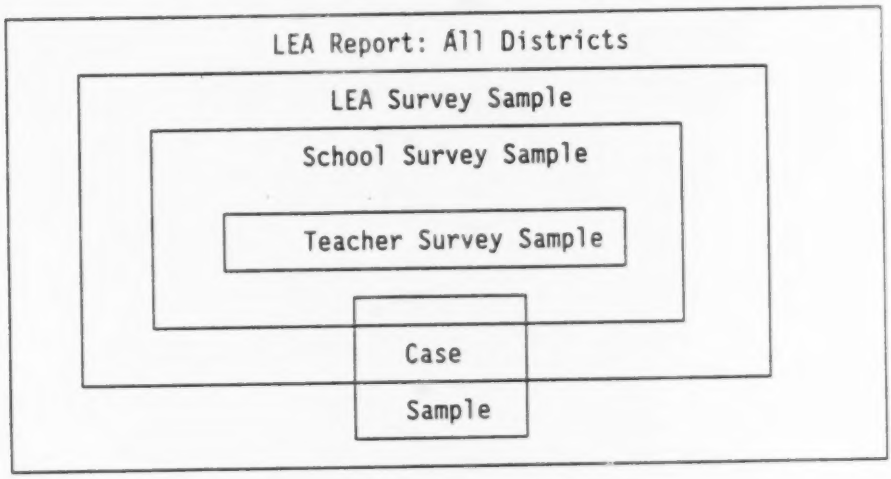
---

* These guidelines are based on practical experience but are not designed to ensure a specific level of statistical precision. We suggest you contact a statistical consultant to advise on the issue of statistical precision.

The sampling technique we have suggested, in which the different subsamples are nested within one another, is graphically portrayed in Figure 8. The School Survey is only administered in districts participating in the LEA survey; the Teacher Survey is only administered in schools which have filed a School Survey. LEAs typically file annual reports. If one were to carry out case studies in selected LEAs, these would not necessarily have to have participated in the survey.

Figure 8

SAMPLING STRATEGY FOR VARIOUS DATA COLLECTION METHODS



The Issue of Survey Response Rates

A survey's response rate is the proportion of sample members (e.g., districts, schools, teachers) included in the sample who actually complete a questionnaire or interview. Because districts (schools, teachers) that do not respond to a survey may be very different from those that do, it is important to use data collection procedures to ensure a high response rate. Fifty respondents from a sample of 200 is not the same as 50 responses from a sample of 75. In the first case, the nonrespondents outnumber respondents by

3 to 1. Even if the nonrespondents were only slightly different from respondents, the overall results would change. By contrast, in the second case, respondents outnumber nonrespondents by 2 to 1. Although there is no magic acceptable return level, it is a good idea to try to obtain 70% or greater response rates. Generally, it is not a good idea to rely on survey results with a response rate of less than 50%. Whatever the response rate, it is a good idea to examine the data you have on nonresponding districts so that you will know the limits of your sample. If there is a very low response rate for one group of districts (e.g., very small districts), you may want to exclude them from your analysis or at least point out that your findings are less representative for this group.

The Issue of Weighting

When all districts have an equal chance of being selected and 100% of them agree to participate in the survey, the sample is self-weighting (each district has a weight of 1). Most commonly, however, some types of districts have a greater chance of being selected (e.g., all large districts are included in the sample) and districts cooperate at different rates (e.g., a greater percentage of medium-size districts respond than large or small districts). In these cases, weighting is needed if the sample as a whole is to be representative. For example, if large districts constitute 50% of the sample, but only 10% of districts in the population of interest, then their responses will need to be multiplied by 1/5th when combined with the responses of other districts.

XI SURVEY DESIGN AND ANALYSIS ISSUES

Throughout this document, we have suggested that SEAs use surveys as one data collection mechanism. Here we offer some practical advice on survey design and use.

Instrument Development

The content of each survey questionnaire should reflect the evaluation questions specified in each analytical framework. For example, the analytical framework for material and equipment acquisition suggests the following topics: what material was acquired or developed, number/percent of teachers who used material, change in what or how teachers teach, change in what students experience, other student outcomes, and teacher outcomes.

You may also want to collect some additional background information, such as, school characteristics (e.g., poverty level, size, "climate"), teacher characteristics (e.g., years teaching, teaching philosophy, attitudes toward computers), and student characteristics (e.g., economic/educational disadvantage, ethnicity). This type of data is very useful for analytical purposes. It allows the analyst to break down the findings by subgroups (high versus low priority schools, for example). It also provides for the opportunity to explain reasons for a program's success or failure (new teachers adopted the material much less quickly than more experienced teachers, for example).

Following are useful guidelines for instrument development:

- You should have a good reason for including items in a questionnaire. Generally speaking, it is preferable to cover a few topics in depth in a questionnaire, rather than to try to cover a wide range of topics. District and school staff are far more likely to complete a questionnaire if it is coherent, does not include items on topics that appear to be irrelevant to the stated purpose of the study, and appears to measure concepts in enough detail to provide meaningful information.

- Items in each instrument should be limited to those for which the target respondent constitutes the best source of information (e.g., Chapter 2 coordinator, principal, teacher, counselor). Thus, for example, information about the goals and purposes of material and equipment acquisition should be collected from the Chapter 2 coordinator. Information about students' experiences and attitudes are more appropriately collected from teachers or the students themselves, if possible. School records and transcripts are the most accurate source for several aspects of students' educational attainments and related behaviors (e.g., attendance, grade level progression).

- To minimize respondent burden and data collection costs, asking factual items of different respondents should be minimized (e.g., the types and cost of materials and equipment purchased with Chapter 2). However, when collecting information on attitudes or perceptions (e.g., the contribution of Chapter 2-supported materials and acquisition) repetition and overlap is desirable (i.e., you will probably want to get the perspectives of district and school administrators, as well as instructional staff).

- Items respondents may consider sensitive should be avoided when possible. Those that are particularly important for addressing a research question should be carefully pretested. We discuss pretesting in more detail later in this document.

Question Format

Data collection instruments usually include a variety of types of questions and a series of statements about which a respondent is asked to indicate agreement or disagreement.

Open-ended versus closed-ended questions. Open-ended questions require respondents to answer in their own words and are rarely used for surveys in which the questionnaire is self-administered (e.g., a mail survey) or in large surveys (e.g., when the sample size is larger than 50). Open-ended questions are more difficult to answer and may be interpreted differently, depending on the perspective of the respondent. For example, the question, "What types of teachers use computer hardware or software supported by Chapter 2?" could produce divergent responses, such as, "teachers of potential dropouts," "elementary school teachers," "math teachers," "teachers in high poverty schools," "new teachers." Furthermore, data processing is more

expensive because responses must be manually coded into meaningful categories before entry onto a computer.

Closed-ended questions require the respondent to select one or more response categories from a list that is provided. Although closed-ended questions provide more uniformity of responses and are easier to process than open-ended questions, they may not capture important responses. To ensure that the categories to a question are exhaustive, it is often a good idea to include an "Other, specify" category. In addition, a "general comments" section can be included at the end of the questionnaire that will allow respondents to provide important information or express opinions that they feel were not captured in the questionnaire. For example, you might ask, "Please provide any other information that will help us to understand better the effect of Chapter 2-supported computer hardware or software on your district's educational program."

Writing questions. In writing questions, the following are useful guidelines:

- Questions should be simple and precise. For example, asking whether Chapter 2-supported computer hardware or software is used by "at risk" students is less clear than asking which of the following types of students use this equipment: Limited English speaking, economically/educationally disadvantaged, handicapped students.

- Instructions should be provided for all questions. For example, respondents may not know whether to circle one or more than one category in response to the question, "What types of students are the principal users of this computer hardware or software?" This can be clarified with a simple instruction to (PLEASE CIRCLE ALL THAT APPLY) or (PLEASE CIRCLE ONE NUMBER).

- Avoid requesting information that requires respondents to refer extensively to records or perform calculations. This is usually frustrating and results in nonresponse. When numerical information is needed (such as dollar amounts or student counts), often estimates can be provided more easily than exact numbers. Let the respondent know that estimates are acceptable, by using the following phrases in the question stem, "about how many teachers," or "approximately what percentage of students." At the end of the question include an instruction to (PLEASE GIVE YOUR BEST ESTIMATE). Even with these instructions, respondents may have difficulty estimating the percentage of teachers or students who use Chapter 2-supported

computer hardware or software. However, they may be able to select from a list of categories, such as: less than 10%, 10% to 25%, 26% to 50%, 51% to 75%, more than 75%.

- Avoid "yes/no" questions, particularly when one response is more socially desirable than the other. For example, asking teachers whether they used the computer hardware or software they were given for their reading program is likely to produce an overwhelming number of "yes" responses. However, asking how much or how often teachers used this equipment (e.g., not very much, some, a lot, a great deal) is likely to produce more meaningful and varied responses.

- Provide respondents with several categories when asking for value judgments or ratings. Usually, 4-point and 6-point scales are used for ratings. For example, you might ask respondents to rate Chapter 2-supported computer software on a 4-point scale, where 1 = poor, 2 = fair, 3 = good, and 4 = excellent. Or you might ask respondents to rate a course they attended on a 6-point scale, where 1 = not at all useful, 2 = not very useful, 3 = somewhat useful, 4 = quite useful, 5 = very useful, and 6 = extremely useful. An even number of categories is preferable to an odd number (e.g., 4 categories rather than 3). This forces respondents to express an opinion (rather than taking the middle category which is usually neutral). A "don't know" or "not sure" category can always be provided for those respondents who are unable to answer the question.

- Avoid asking respondents to rank order their responses. These types of items tend to be difficult and time-consuming. (Often respondents have to reread the response categories several times.) It is simpler to ask respondents to indicate all the categories that apply and then the category that applies most. For example, you might ask respondents to indicate all the benefits of Chapter-2 supported materials and equipment, and then ask how they most benefited students.

- Ask questions that will allow respondents to express both positive and negative sentiments. For example, rather than just asking teachers to rate a Chapter 2-supported inservice training on peer tutoring overall, you might ask them to rate it on a variety of dimensions (e.g., whether the material covered was new, informative, useful). You may want to ask respondents how this course compared with other inservice training in which they have participated (e.g., much more useful, somewhat more useful, about the same, less useful, much less useful). You will probably want to know to what extent respondents have tried to use what they learned, and whether they had problems. Also, you can ask in what ways the inservice training could be improved. This series of questions will allow respondents to express a variety of opinions, both positive and negative.

- Agree/disagree statements should be worded so that agreement does not always indicate a positive response because some respondents tend to agree with almost everything ("yea-sayers") and others tend to disagree ("nay-sayers"). For example, you might ask the extent to which a teacher agrees or disagrees with the following statements

about the inservice training course they attended: "The course provided me with new information"; "The information was not clearly presented"; "The course did not provide enough practical applications for the types of students I teach"; "I would recommend that other teachers take this course." The following 4-point scale is often used for agree/disagree items: "strongly disagree, somewhat disagree, somewhat agree, and strongly agree."

Questionnaire Length

Shorter questionnaires do not necessarily produce higher response rates than longer ones. The relevance of the subject matter and the ease of completing the questionnaire is far more important to respondents. Nevertheless, there are "typical" interview lengths, depending on the type of data collection method used:

- Mail questionnaires for district and school surveys tend to range from a few pages up to about 20. Depending on the type and length of the question, you can ask 4 to 8 closed-ended questions per page.
- Telephone surveys last 15 to 20 minutes, on average. Again, depending on the length of the question (and the amount of time it takes for a respondent to answer the question), you can ask from 1 to 3 closed-ended questions per minute.
- Personal interviews usually last from 30 minutes to 1 hour. Because personal interviews usually include some open-ended questions, you usually ask fewer questions than you would in the same amount of time in a telephone interview.

Use of skip instructions. Often, questionnaires are designed in separate sections to capture a wide range of activities and functions that can occur. For example, for Chapter 2 a questionnaire might be divided into separate sections on materials and equipment acquisition, professional development, and direct services to students. Depending on the district or school, some sections may not apply. To minimize the amount of time required to complete the questionnaire, skip instructions are provided. For example, the first question in Section A (Chapter 2 Instructional Resource Support), might ask, "In the current school year, are any Chapter 2 funds being used to purchase materials or equipment for public school libraries, media centers, or other departments?" If the answer is yes, the respondent is instructed to

(PLEASE ANSWER Q. A2). If the answer is no, the respondent is instructed to (PLEASE SKIP TO SECTION B, PAGE 4). Using this approach, you can point out in the cover letter (or in the introduction to the interview) that most districts or schools will complete only one or two sections of the questionnaire, and therefore, the average time it will take to complete the questionnaire is fairly short.

Pretesting

It is always a good idea to try out each questionnaire on a small sample (5 to 9 respondents per questionnaire) before the survey is fielded. The sample for a pretest is usually purposively selected to represent the range of potential respondents for each questionnaire (e.g., Chapter 2 coordinators in large urban districts, medium-size urban districts, small rural districts, and so on).

The pretest involves having the respondent complete the questionnaire in the same manner as you plan to use in the full study (e.g., mail questionnaires are pretested by mail, telephone questionnaires by telephone and so on). In addition, during pretests of telephone and personal interview questionnaires, interviewers record their experiences immediately following each interview. Also, they make notations in the questionnaires to record pretest respondents' comments, as well as any unusual occurrences related to administration. Regardless of the interviewing method (e.g., mail, telephone, personal interview), all pretest respondents (including mail survey respondents) should be interviewed about the following issues after completing the questionnaire: ease or difficulty of completing the questionnaire; why items were left blank/not answered; why instructions were incorrectly followed; appropriateness of wording and response categories, particularly for important items; length of time to complete the questionnaire; whether any records were consulted or help was required from another person to complete the interview. Because sometimes it is desirable to debrief pretest respondents in-person (so that you can look over the questionnaire together), pretests often are conducted locally.

After the pretest, interviewers should meet as a group, review each questionnaire, and thoroughly discuss items with which interviewers or respondents had problems. Other topics that can be discussed include problems with layout or questionnaire design, questions or concerns expressed by respondents, problems with obtaining cooperation, and the interview setting.

Respondent Identification

In surveys of district and school staff, often you will know the type of person you want to interview but not his or her name or title (e.g., someone knowledgeable about Chapter 2, someone who is using computers purchased by Chapter 2 in their reading program). For example, depending on the size of the district, the person who is most knowledgeable about Chapter 2-supported programs and purchases can be an assistant superintendent for curriculum and instruction, the state and federal programs coordinator, the Chapter 2 coordinator, or a teacher. To identify the appropriate respondent, a brief phone call should be made to the district or school in advance of mailing a questionnaire or trying to schedule an interview. Questionnaires should always be mailed to a specific individual and not a person with a particular title (e.g., Chapter 2 coordinator). Also, although you may want to contact the district superintendent or principal to identify the appropriate respondent, do not ask him/her to give the questionnaire to this person. Personally addressed questionnaires are far more likely to be completed and returned, although, as we mentioned in the previous section on sampling, it may be impossible to identify teachers by name. Finally, if the questionnaire is not returned in a timely manner, you will want to follow up directly with the appropriate person (either by mail or phone).

Data Collection

Surveys can be conducted in-person, by phone, or by mail. Below we highlight some of the advantages and disadvantages of each data collection method:

- Personal interviews. Unless large numbers of persons are going to be interviewed at the same time and location, personal interviews tend to be very costly because of the need to pay for travel time and expenses. Also, often personal interviews include some open-ended questions and therefore, they need to be conducted by more highly skilled staff than more structured telephone interviews. The advantage of in-person interviews is that they can be longer than telephone interviews. If during the middle of the interview, you determine that you are speaking with the wrong person, you can arrange an interview with the appropriate person. It is easier to ask open-ended questions (because you can clarify your meaning), and you may be able to consult district or school records to clarify interview results. For complex questions and questions with a large number of response categories, you can give respondents visual aids, such as lists.

- Telephone interviews. Telephone interviews tend to be cheaper than personal interviews, but more expensive than mail surveys. A large number of telephone interviews with district and school administrators can be conducted in a relatively short time-frame (e.g., a month). Like personal interviews, if you determine you are talking to the wrong person, you can terminate the interview and reschedule an interview with the appropriate person. You can provide additional instructions if the respondent does not understand the question. The disadvantage of telephone surveys is that the questions need to be very simple, and you will want to limit the number of response categories (to 6 or less) so that they will be easy for respondents to remember. Furthermore, telephone surveys do not work well for instructional staff. Often, teachers do not have telephones in their classroom or have a limited amount of time during the day in which they can be interviewed. (You probably will not have access to teachers' home phone numbers, either.)

- Mail surveys. Mail questionnaires are cheaper than telephone or personal interviews. High response rates can be obtained, but you will have to do several follow-up mailings and possibly, reminder phone calls to accomplish this. The advantage of mail questionnaires is that they can be longer and include more complex questions than telephone interviews. Also, respondents can complete them in the privacy of their home or office, and therefore, they appear to be more confidential. The disadvantage of mail surveys is that they are difficult to conduct in a short time frame (less than 2 months), because it is left to the respondent to decide when to complete the questionnaire. Also, you may not know the conditions under which the questionnaire was completed (e.g., did the school principal and teacher jointly complete the questionnaire or did the teacher complete it on his or her own?).

Ensuring High Response Rates

Whatever the survey methodology, it is important to obtain a high response rate to minimize bias (see discussion in previous section). A variety of methods can be used to accomplish this.

- Notify district superintendents and school principals. It is very important to notify the district superintendent and school principal (for schools in which you will be interviewing teachers) that you are conducting a survey and provide them with a phone number that they can use to obtain more information about the study.

- Send advance letters. Mailing a letter to sample members explaining the purpose and background of the study in advance of mailing the questionnaire or arranging a telephone or personal interview can be used to enhance response rates. In this mailing you can indicate that the district superintendent and school principal have been informed about the study. Also, you should stress the importance of all respondents' participation even if their district or school receives only a small amount of Chapter 2 funds

- Ensure respondent confidentiality. In most surveys, questionnaires are not anonymous: an identification number is usually assigned to each individual and this is placed on the cover of the questionnaire in order to link the respondent to a school and district and to identify nonrespondents for follow-up mailings. However, responses to questionnaires can be kept confidential by not putting any individual's name or address any place on the questionnaire or including it in the same data file that includes questionnaire responses. Responses for individual teachers, administrators, schools or districts should not be reported. Instead, responses should be reported for "groups" of respondents. For example, you can describe the results for large districts separately from medium-size districts, and small districts. In states in which one or two districts are very large, it may be difficult to protect the confidentiality of those districts. Nevertheless, assurances of confidentiality are important in ensuring a high response rate and in obtaining "truthful" responses.

- Conduct follow-ups. For telephone or in-person interviews, it is important to make several attempts on different days and at different times to conduct or schedule an interview. (For telephone surveys, usually a response rate of 50% can be achieved with four contact attempts.) The interviewing period should be sufficiently long (at least a month) to accommodate sample members' vacations, illnesses, and busy schedules. For mail questionnaires, several different types of follow-ups can be conducted. Usually, a reminder postcard is mailed approximately 10 days to 2 weeks after the initial copy of the questionnaire is distributed. About a month after the initial mailing, a second copy of the questionnaire is mailed to

nonrespondents. All mailings should be sent first class. Otherwise, your reminder postcard might arrive before the initial copy of the questionnaire. Usually two follow-up mailings are sufficient to obtain a moderate to high response rate (50% to 75%). If a higher response rate is needed, nonrespondents can be contacted by telephone, and a third copy of the questionnaire can be mailed to those who request one. Finally, nonrespondents can be interviewed by telephone.

- Send a complete package. For mail questionnaires it is important to send a package that includes a cover letter (similar to the advance mailing) that explains the purposes of the study and that includes a phone number that respondents can call if they have any questions. Also, the questionnaire should include a self-addressed postage-paid (or business reply) envelope that respondents can use to return their completed questionnaire. (Because these envelopes sometimes get misplaced, be sure to include a return address on the questionnaire.)

- Provide a copy of the study's results. Usually respondents will be more likely to participate in a study if they believe they are going to get something out of it. Although you will not need to send a copy of the study's results to every respondent, you can indicate that you will be sending a copy to the district superintendent or Chapter 2 coordinator when the study is completed.

Survey Costs

Survey costs are a function of the type of data collection method (i.e., in-person interviews are the most expensive, whereas mail surveys are the least), the overall sample size, the length of the interview (particularly for telephone surveys), and the type and amount of analysis you plan to perform. Because surveys tend to be costly, it is usually preferable to conduct a well-designed survey on a few topics with a small representative sample for which you obtain a high response rate than a survey on a variety of topics with a large sample.

After you have designed the questionnaires, it is often cheaper to contract with a survey firm to collect the data, especially for mail and telephone surveys. For telephone interviews, this is only a good idea when the questionnaire is structured and does not require highly skilled interviewers. Because in-person interviews tend to be semi-structured or include open-ended questions, it is usually not a good idea to subcontract with data collection firms for these types of surveys.

Timing Considerations

Timing is an important factor in obtaining a high response rate for surveys. Even for district surveys, it is not a good idea to attempt to collect data during the summer. District and school surveys are best conducted in the early fall (beginning the middle of September) so that you can be out of the field before Thanksgiving or early spring (January through May). In planning data collection, be sure to take into account the time required for advance mailings or telephone contacts, distributing the questionnaire or scheduling interviews, and follow-up activities.

Data Processing

Even for very small samples (e.g., samples of 25 or less) and simple analyses (e.g., frequencies, cross-tabulations) it is usually much easier (and more accurate) to tabulate survey results on a computer using a spreadsheet program (e.g., Lotus) or a statistical software package (e.g., Statistical Package for the Social Sciences, SAS) than to do this manually.

Regardless of whether you use a computer, responses to each questionnaire should be reviewed for internal consistency before they are tabulated (e.g., NUMBER OF STUDENTS IN CHAPTER 2-SUPPORTED PROGRAMS $\leq$ TOTAL SCHOOL ENROLLMENT). This review should also be designed to ensure that numerical and other open-ended responses assume only meaningful values (e.g., if RESPONDENT'S SEX = M or F, then SEX = A is not permissible) and that values are not outside normal bounds (e.g., if AGE OF STUDENTS SERVED > 22). If responses to key items do not "make sense," you may want to recontact the respondent by telephone to clarify the response. If this is too expensive, you may want to exclude these questionnaires (or items) from your analyses.

Measurement Construction

Often it is desirable to combine responses to several items on the same topic to create a single measure, particularly for attitudes or ratings. This will simplify the analysis and probably will be more reliable than

analyses of responses to individual items that are extremely sensitive to question wording. Composite measures can be complex and can be constructed using a variety of statistical techniques. However, most of the time, composite measures are created by merely adding and averaging scores assigned to sets of items. For example, a teacher might be asked to rate a set of instructional materials on four characteristics (e.g., appropriateness for students' intellectual level, relevance to the curriculum, interest level, ease of use) on a 4-point scale (where 1 = poor, 2 = fair, 3 = good, and 4 = excellent). Responses to these four items could then be added and averaged to obtain an overall rating of these instructional materials.

Analysis

Most analyses of survey data begin by calculating the percentage of respondents who selected each response category to a question (i.e., the percentage of districts that said "yes" to Question 1). A next step is to pick three or four characteristics of districts or schools that you feel are important in understanding the survey results (e.g. enrollment, poverty level) and to form subgroups based on each of these characteristics. You can then calculate and compare the percentage of respondents for each subgroup who selected each response category (e.g., the percentage of large districts that said "yes" is greater than the percentage of "small" districts). Because the reliability of subgroup analyses is based on sample size, you should not divide the sample into very small subgroups. (For small samples, often large percentage differences are not statistically significant). Thus, if your overall sample is small, you may only be able to compare large and small districts. However, if your sample is relatively large, you may be able to compare more subgroups (e.g., large, medium-size, and small districts).

If you are concerned about the statistical reliability of subgroup differences or want to perform more complicated survey analyses, you should consult with a statistician. If your data are entered into a computer, using a statistical software package, you can perform some simple statistical tests with a small amount of consultation. More important than statistical

analyses is that the survey results are "logical" or "make sense." If they do not, you will need to reexamine the methods you used to conduct the survey (e.g., how was the sample selected, what types of districts/schools responded, is it possible that survey respondents misinterpreted the question).

XII CASE STUDIES

Case studies are relatively in-depth examinations of a specific activity or program (e.g., drop out prevention programs) in a single or small set of instances (usually less then 20). In the field of educational evaluation, case studies usually involve on-site data collection, including interviews, record reviews, and--sometimes--participant observation.

Case studies have two main shortcomings: they are extremely costly and collect data from fewer cases than most other methods. In some instances, however, case studies are the only plausible way of studying a particular program or set of activities; and in many instances, case studies are the preferable method because of the richness of data they can provide. In particular, case studies are useful when:

- Significant resources have been invested in demonstration projects and the SEA wants detailed information on the success of the projects.
- There is a need to understand well "why" a certain intervention failed or succeeded so that these lessons can be applied to other districts and schools interested in adopting a similar project.
- The activity under study is so innovative or complex that the evaluator can only create a loose evaluation design before examining the activity in more detail.

Sampling Cases

Because the sample of districts in case studies is usually small (ranging from a few districts up to 20), a simple or systematic random sample is unlikely to provide enough contrast on background characteristics (e.g., size, poverty level) or how Chapter 2 resources are used to support a good case study analysis. (A small random sample is unlikely to include types of districts that represent a small percentage of all districts, such as large districts or those that use Chapter 2 funds differently from the majority of

districts.) For this reason, district samples (and samples of schools within districts) for case studies are usually purposively selected. Also, because visits to districts for case studies involve travel time and expenses, the sample is usually geographically clustered (i.e., several districts that are within an hour driving distance of each other are selected in each region of the state).

Sample selection begins by developing a list of candidate districts that vary on key dimensions (e.g., background characteristics, uses of Chapter 2). These districts can be identified by reviewing state Chapter 2 applications, from other state databases on districts, and by contacting a small sample of districts by telephone and other persons knowledgeable in the state (e.g., educational associations, researchers) for nominations.

Selection of schools usually occurs after districts have been notified and agreed to participate in the case study. Because usually only two or three schools are visited in each district, these are also purposively selected (e.g., to represent variation in grade levels served and uses of Chapter 2.)

Instrument Development

Unlike questionnaires for large sample surveys, instruments for case studies are usually unstructured or semi-structured. Often case study protocols consist of a list of topics, subtopics, and probes designed to elicit examples from respondents rather than general statements. Although the same topics are pursued with different types of respondents, often they are stated differently. Space can be provided on the topical guides for interviewers to record interview responses.

Because the topics to be pursued will differ somewhat based on who is being interviewed and their role, usually more than one topical guide is developed. For example, one topical guide might be developed for each of the following types of respondents: top district administrators (e.g., superintendent, assistant superintendent), Chapter 2 coordinator, other district

administrator, school-level administrator, and instructional staff. Interviews with the superintendent and school principal can be used to obtain background information (e.g., district/school characteristics, priorities, problems), whereas interviews with other administrators and instructional staff should focus on Chapter 2-supported programs and purchases.

If you plan to review student records (e.g., for grades, test scores, progress through core curriculum, graduation status), it is a good idea to prepare a more structured instrument (similar to a closed-ended survey questionnaire) for this purpose.

Preparing for Data Collection

Following are guidelines for preparing for visits to case study districts:

- Pilot test. A pilot test of data collection instruments and procedures should be conducted in one district before training interviewers and scheduling other district visits. The pilot test will help you to determine the number and types of respondents who will need to be interviewed to obtain the desired information, refine the topics to be covered with each respondent, and estimate the amount of time needed for individual interviews and for the district visit.

- Requirements for district/school visitors. Because instruments for site visits usually are unstructured, site visitors will need to be knowledgeable about educational programs in the state and about Chapter 2. However, different staff from those who normally conduct monitoring or compliance visits should be used to conduct the case studies. This will help to set site visit respondents at ease and to assure them that their responses will be confidential.

- Training sessions. The purposes of training sessions are to ensure that site visitors have a common understanding of the purposes of the study and that meaningful data are collected. Depending on the background of site visitors and the complexity of the study, a training session can last between one and three days. The training session should provide an overview of the study and study design and item-by-item training on the interviews, including probing for concrete examples.

- Initial district contact. First, a letter should be mailed to the superintendent explaining the purposes of the study and that you would like to visit the district (include a telephone number that the

superintendent can use to call for more information). About 1 week after the letter is mailed, the person who is going to visit the district should contact the superintendent by telephone to answer questions and try to stimulate interest in the study's goals. Assure the superintendent that this is not a monitoring or compliance visit. After the superintendent has agreed to participate in the study, it is a good idea to have him or her designate a person with whom you can schedule or coordinate your visit. (This will probably be the person responsible for Chapter 2.)

- Scheduling interviews. You will want to discuss with the district's site visit coordinator (e.g., the superintendent's designee) appropriate district respondents. Also, you will want to identify schools to visit. Once these are chosen, you can contact principals to identify appropriate school level respondents. Most interviews should be scheduled before you arrive at the district. This will enable you to use your time more efficiently on site. Scheduling interviews will require several phone calls. In some cases a district staff or school person may offer to schedule the interviews for you. In those instances, be sure to review the schedule with this person before you arrive at the district. It is always a good idea to confirm the appointments that have been arranged for you (either by phone or mail) before you arrive. Allow a half hour between scheduled appointments. This will give you flexibility if the interview starts late, gets interrupted, or runs longer than anticipated. Also, allow for some unscheduled time so that you can meet with relevant respondents who are identified during the course of interviews.

- Collect and review background information. Interviews will be more efficient and informative if you collect and review background information on the district (e.g., Chapter 2 application, brochures) and school (e.g., descriptions of school programs) before your site visit. It will also help to establish rapport with the persons you interview if you are well prepared.

Conducting District Visits

Following are some guidelines for conducting district visits:

- Use two-person teams to conduct district visits, if possible. Two-person teams offer more flexibility to the district/school in scheduling interviews. Also, a team of interviewers can complete a district visit in less time than one interviewer; this decreases the burden to the district/school.

- Major analysts and report writers should conduct some of the district visits. This will add to their base of knowledge and help them to understand better the data gathered by other district visitors.

- Using a two-person team, district visits can be completed in two to three days, depending on the size of the district, the number of schools that will be visited, the number of persons to be interviewed, and the amount of information to be collected. A two-person team might spend 1 day at the district together, and each member visit a different school on the second day.

- Site visit teams should conduct background interviews and other key interviews together (e.g., with the superintendent, Chapter 2 coordinator) so that they can share their interpretation of these interviews and refine the topics to be covered with other respondents. Although most other interviews should be conducted separately (to be cost-effective), site visit teams should continue to share experiences while on site to resolve discrepancies and ensure that all of the necessary information is collected.

- With good scheduling, a two-person team can interview from eight to twelve persons per day on site. However, do not overschedule. Although multiple perspectives are desirable, the value of a site visit comes from the amount and depth of information collected, not from the number of persons interviewed.

- Student record reviews require a considerable amount of time to perform (at least a half hour per record). If you are going to review student records, be sure that you advise district and school staff beforehand. If possible, student records should be located and waiting for you when you arrive. Also, you should arrange for a district or school staff member to help you perform the first two or three record reviews to make sure you understand how the records are organized and recording conventions.

Data Analysis and Reporting

When district visits are completed, several types of data will be available for each site (e.g., background information, documents, field notes, staff questionnaire, student record forms). A reporting form should be prepared that interviewers can use to organize and record their field notes. These forms are usually organized by topic (e.g., background information, materials and equipment acquisition, professional development, direct services to students, teacher outcomes, student outcomes). The reporting forms are designed to integrate all of the information collected about a single site and to facilitate comparisons of different districts.

Analysis of findings from different districts is usually done through group discussions in which all site visitors participate and by using the reporting forms to compare districts. Analysis involves grouping districts in different ways (e.g., by size, poverty level, other student characteristics) and then comparing their use and experiences with Chapter 2.

District and school names should not be used in case study reports. Findings can be reported for groups of similar districts (e.g., large districts tended to use Chapter 2 to support computer purchases). Sometimes in a case study it is desirable to describe individual districts. In those instances, the name of the district can be changed and contextual information can be provided (but not enough to identify the district)--e.g., District A, a small rural district with a high concentration of limited English proficient students uses Chapter 2 to support its bilingual program.

XIII SELECTED BIBLIOGRAPHY OF RESOURCES

Babbie, E. R. (1973). Survey research methods. Belmont, CA: Wadsworth.

Cronbach, L. J., Robinson Ambron, S., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., Walker, D. F., & Weiner, S. S. (1980). Toward reform of program evaluation. San Francisco: Jossey-Bass.

Erdos, P. L. (1983). Professional mail surveys. Malabar, FL: Robert E. Krieger.

Groves, R. M., & Kahn, R. L. (1979). Surveys by telephone - a national comparison with personal interviews. San Francisco: Academic Press.

Kennedy, M. M. (1979). Generalizing from single case studies. Evaluation Quarterly, (3)4, 661-678.

Kish, L. (1965). Survey sampling. New York: John Wiley & Sons.

Labaw, P. J. (1980). Advanced questionnaire design. Cambridge, MA: Abt Books.

Rossi, P. H., Wright, J. D., & Anderson, A. B. (Eds.). (1983). Handbook of survey research. San Francisco: Academic Press.

Sudman, S., & Bradburn, N. M. (1982). Asking questions - a practical guide to questionnaire design. San Francisco: Jossey-Bass.

Yin, R. K. (1984). Case study research design and methods. Beverly Hills, CA: SAGE.